Joanna Aho

The Effect of Visual Attention on Spatial Processing in Human Auditory Cortex Studied with Magnetoencephalography

Faculty of Electronics, Communications and Automation

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo, 10 May 2010

Thesis supervisor:

Prof. Mikko Sams

Thesis instructor:

M.A. Nelli Salminen

Aalto University
School of Science
and Technology

AALTO UNIVERSITY
SCHOOL OF SCIENCE AND TECHNOLOGY

ABSTRACT OF THE
MASTER'S THESIS

Author: Joanna Aho

Title: The Effect of Visual Attention on Spatial Processing in Human Auditory Cortex Studied with Magnetoencephalography

Date: 10 May 2010 Language: English Number of pages: 7+54

Faculty of Electronics, Communications and Automation

Department of Biomedical Engineering and Computational Science

Professorship: Cognitive Technology Code: S-114

Supervisor: Prof. Mikko Sams

Instructor: M.A. Nelli Salminen

Previous studies on spatial processing in the human auditory cortex have found evidence that sound source locations are coded with the relative activity of two opponent populations of neurons: one tuned to the left and the other to the right hemifield. However, little is known about how visual spatial attention modulates this code. In this study, the subjects were asked to attend to visual stimuli that were presented at either spatially congruent or incongruent locations with the auditory stimuli. The brain responses to the auditory stimuli were measured with magnetoencephalography (MEG). It was found that regardless of the direction of the attention, attending to the visual stimuli equally enhanced the brain responses elicited by auditory stimuli compared to the condition requiring no attention. The results of this thesis suggest that spatially attending to visual stimuli increases the overall sensitivity of spatially selective auditory cortical neurons.

Keywords: visual spatial attention, auditory spatial processing, human auditory cortex, MEG, stimulus-specific adaptation paradigm, population rate code

AALTO-YLIOPISTO
TEKNILLINEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Joanna Aho

Työn nimi: Magnetoenkefalografiatutkimus visuaalisen tarkkaavaisuuden vaikutuksesta äänen tulosuunnan prosessointiin ihmisen kuuloaivokuorella

Päivämäärä: 10.5.2010 Kieli: Englanti Sivumäärä: 7+54

Elektroniikan, tietoliikenteen ja automaation tiedekunta

Lääketieteellisen tekniikan ja laskennallisen tieteen laitos

Professuuri: Kognitiivinen teknologia Koodi: S-114

Valvoja: Prof. Mikko Sams

Ohjaaja: FM Nelli Salminen

Aiemmissa ihmisen kuuloaivokuorta koskevissa tutkimuksissa on löydetty viitteitä siitä, että äänilähteen sijainnin koodaamiseen osallistuu kaksi kuuloaivokuoren kilpailevaa hermosolupopulaatiota. Näistä toinen on virittynyt reagoimaan vasemmalta ja toinen oikealta tuleviin ääniin. Ei kuitenkaan tiedetä, miten tarkkaavaisuuden suuntaaminen visuaalisiin ärsykkeisiin vaikuttaa äänilähteen paikan koodaamiseen. Tässä tutkimuksessa koehenkilöitä pyydettiin tarkkailemaan visuaalisia ärsykkeitä, jotka esitettiin joko samasta tai eri suunnasta kuin ääniärsykkeet. Ääniärsykkeiden herättämiä vasteita mitattiin kuuloaivokuorelta aivomagneettikäyrän (MEG) avulla. Tutkimuksessa huomattiin, että visuaalisella tarkkaavaisuudella oli tarkkaavaisuuden suunnasta riippumaton voimistava vaikutus äänten aiheuttamiin aivovasteisiin verrattuna tilanteeseen, jossa tarkkaavaisuutta ei vaadittu. Tämän diplomityön tulokset viittaavat siihen, että tarkkaavaisuuden suuntaaminen visuaalisiin ärsykkeisiin herkistää yleisesti äänen tulosuunnan koodaamiseen osallistuvia hermosoluja.

Avainsanat: visuaalinen tarkkaavaisuus, suuntakuulo, ihmisen kuuloaivokuori, MEG, adaptaatioparadigma

Preface

Without the help and support from the following people this thesis would never have been possible.

First, I would like to thank the Faculty of Electronics, Communications and Automation for the scholarship that allowed me to focus on this thesis. This meant a lot to me.

I would like to thank Professor Mikko Sams, my supervisor, for giving me the opportunity to work on this fascinating project in his group.

I wish to express my sincere appreciation and warmest thanks to my instructor M.A. Nelli Salminen for all her technical and intellectual support during this project. She has always been there to patiently answer any questions that I may have had. Her encouragement, kindness and positive attitude at all stages of this project have carried me to this point. I feel words are not enough to thank her.

I want to thank Mia Illman from the Low Temperature Laboratory for kindly helping me with any sudden technical problems I encountered with the MEG equipment. I also want to thank–in no particular order–Kaija Virolainen, Juha Salmitaival, Timo Aarnio and Mikko Hakala from the Department of Biomedical Engineering and Computational Science for helping me with practical issues during this project.

Many thanks to the friends and strangers who made this thesis possible by being my fearless subjects.

Special thanks to my friend and fellow student Antti Kunnas for encouraging me to contact M.A. Nelli Salminen, who ended up being my instructor. If it hadn't been for this lucky coincidence, I might not be writing this preface now.

I would like to thank my faraway friends Dino and Wanda for reading my thesis in the early stages and for providing valuable feedback. They kept me sane and awake during the long hours at night that I spent on writing this thesis.

I would also like to thank Laura for her invaluable friendship. She has always shown interest in what I do and pushed me forward ever since we met.

This thesis is dedicated to my loving family: Mom, Dad, Jukka, Mika and Hanne. Thank you for your lifelong love and support, and thank you for always being there for me. You never stopped believing in me.

Otaniemi, 10 May 2010

Joanna Aho

Contents

Symbols and abbreviations

Symbols

θ	azimuth, horizontal angle
δ	elevation, vertical angle
$\widehat{=}$	corresponds to
$\overline{x}$	arithmetic mean (average)
σ	standard deviation

Abbreviations

AEF	auditory-evoked field
AEP	auditory-evoked potential
ANOVA	analysis of variance
EEG	electroencephalography
EOG	electrooculogram
ERF	event-related field
ERP	event-related potential
fMRI	functional magnetic resonance imaging
HPI	head position indicator
HRTF	head-related transfer function
ILD	interaural level difference
ISI	interstimulus interval
ITD	interaural time difference
MEG	magnetoencephalography
N1	the largest negative deviation of the ERP
N1m	the magnetic counterpart of the N1
Nd(m)	negative difference
PET	positron emission tomography
PN(m)	processing negativity
SNR	signal-to-noise ratio
SQUID	superconducting quantum interference device

1 Introduction

Compared to vision and touch, the processing of spatial information is not a trivial task for the human auditory system. In vision, the image of the world is depicted on the retina, and the spatial relations of the objects that we are looking at are readily available for the neurons connected to the retina. Similarly to vision, the spatial information of an object causing, for example, pain in the skin is available for the neurons having touch receptors in the painful area. For audition, however, the neurons in the inner ear are specialized in processing the frequencies of sounds rather than determining the direction of a sound source. The ability to determine the location of a sound source is, therefore, a result of complex computations based on frequencies, and the intensity and the time differences between the two ears.

Previous studies on spatial processing in human and primate auditory cortices have found evidence that sound source locations are coded with the relative activity of two opponent populations of neurons: one activated by sounds coming from the left and the other by those coming from the right hemifield [1, 2]. This is known as the *population rate code of auditory space* [2]. The code was first found in the monkey auditory cortex by using invasive brain imaging methods not applicable in human subjects. Until recent years, there had been very few findings which would have proved that the code actually exists also in the human auditory cortex. It has been proposed that the differences in the level of activation between the two populations of neurons are so small that the non-invasive brain imaging methods are not able to detect them [1]. Proving the existence of the code has been challenging also due to the fact that the auditory cortical neurons that code the auditory space seem to be intermingled rather than clustered together. Current non-invasive brain imaging methods are not able to detect the activity of individual neurons but rather the combined activity of a group of neurons.

In a recent magnetoencephalography study [2], the existence of the population rate code in human auditory cortex was successfully revealed by using the *stimulus-specific adaptation paradigm*, introduced by Robert A. Butler in 1968 [3]. In this paradigm, the level of neuronal adaptation to repeated sounds is used as an indicator of how much the neuronal populations activated by the sounds overlap in the cortex. It was found that two sounds, presented repeatedly in different locations, caused less adaptation to the neural responses than sounds presented in the same location, suggesting that the auditory cortical neurons are selective to different sound locations [2].

Selective attention is a cognitive process that allows us to focus on relevant inputs and ignore irrelevant ones. Selective attention can be either *overt* or *covert*. In overt attention the sense organs (e.g., eyes) are directed toward the attended stimulus. Covert attention, in contrast, means the ability of mentally focusing on a relevant stimulus without changing the position of the sense organs.

It is well known that selective attention increases the neural activity in cortical areas related to the attended modality (e.g., auditory, or visual cortices). Studies on dichotic listening and auditory spatial attention have reported that selective attention has unimodal effects, enhancing the brain responses to the attended stimulus

type. In dichotic listening experiments the subject was typically presented with independent auditory streams simultaneously at both ears and asked to attend to deviant sounds at one ear only (e.g., [4, 5]). It was found that the attended stimuli received enhanced processing. Similar results were found in experiments studying auditory spatial attention, suggesting that the stimuli presented at attended locations elicit larger responses than those occurring at unattended locations (e.g., [6]).

Selective attention can also have cross-modal (involving more than one sense) effects on the processing of unattended auditory stimuli. Previous studies have provided evidence that directing attention to a specific location within vision or touch increases the brain responses to the simultaneously presented auditory stimuli occurring at that location, even if the auditory modality was unattended [7, 8]. Contradicting findings have also been found, showing that the activity of the auditory cortical neurons decreased when attention was directed toward another modality [9, 10]. However, results of the suppressive effects of selective attention are controversial, because the referenced experiments did not use auditory stimuli to assess the unattended auditory modality.

The primary interest of this thesis was to investigate using the magnetoencephalography whether visual spatial attention can modulate the neural coding of sound source location (the population rate code) in the human auditory cortex. Based on the knowledge from previous studies, the stimulus-specific adaptation paradigm was chosen to reveal the population rate code of auditory space. To see the possible modulation effects of visual spatial attention on this code, the subjects were given a visual task that would direct their attention either to the left or to the right side. In order to determine whether visually attending to the location of the sound increases the brain responses elicited by the sounds, the subjects were presented with task-irrelevant sounds that were asynchronous with the visual stimuli. The subjects were, thus, performing the visual task either in the same or in the opposite side as the sounds while asked to ignore the sounds. The magnetic brain responses elicited by the sounds were measured from the auditory cortical areas to determine whether attending to the visual task modulates the population rate code of auditory space.

2 Auditory Space Perception

The auditory system plays an important role for both humans and animals not only in communication but also in localizing sound sources that surround us. The ability to localize the direction from where a sound is coming helps us in seeking and avoiding objects even when they are out of sight. It also helps us direct our visual attention toward the appropriate direction. Human auditory system is based on complex mechanisms and is capable of localizing sounds surprisingly accurately, providing valuable information about the locations of objects and events. To give insight into the mechanisms that the human auditory system employs to determine the direction of a sound, the basics of sound localization will be explained first.

2.1 Coordinates for Sound Source Localization

The directions of sound sources in space are often defined by three planes that intersect each other in the center of the head as illustrated in Figure 1 [11, 12]. The *horizontal plane* can be defined roughly by the two ear canals and the tip of the nose. The *median plane* bisects the head vertically through the nose and divides the head in two equal parts (left and right). The *frontal plane* is perpendicular to the previous two planes, and it intersects the entrances to the ear canals, dividing the head vertically into anterior and posterior sections.

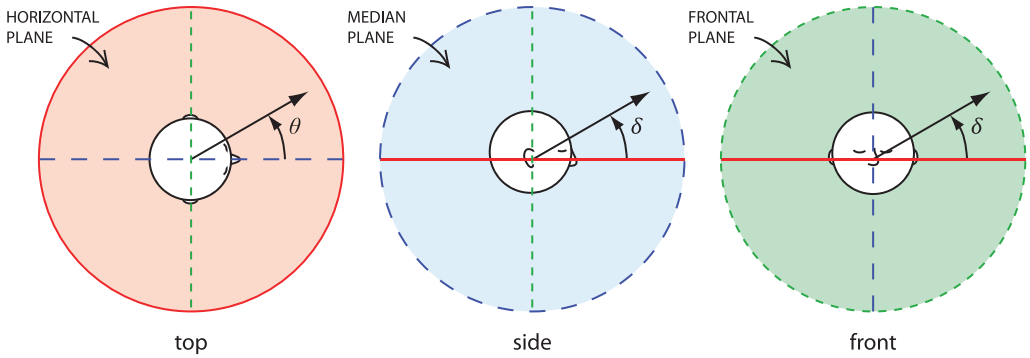

Figure 1: The direction of a sound source is defined by three planes (horizontal, median and frontal) that are perpendicular to each other. The horizontal plane (solid red line), intersects the entrances to the ear canals and divides the head horizontally into upper and lower parts. The median plane (dashed blue line) bisects the head vertically through the nose and divides the head in two equal parts. The frontal plane (dashed green line) intersects the entrances to the ear canals, dividing the head vertically into anterior and posterior sections. In addition to the planes, two angular measures are defined: azimuth θ (the horizontal angle) and elevation δ (the vertical angle).

The direction of a sound source can now be specified by two angular measures: *azimuth* θ and *elevation* δ. Azimuth, or the horizontal angle, is the angle produced by the projection of a sound source onto the horizontal plane. It describes in degrees (0–180°) how much the sound source has rotated around the listener's head. Azimuth is 0° in front of the listener and 180° behind the listener. The angle increases negatively for leftward directions and positively for rightward directions.

Elevation, or the vertical angle, is the angle produced by the projection of a sound source onto the median plane. In other words, it is the angle between the sound source and the horizontal plane. It describes how many degrees (0–90°) the sound source is above or below the horizontal plane. Elevation is positive for sound sources that lie above the horizontal plane and negative for those below the horizontal plane. For example, a sound source with 0° azimuth and 0° elevation lies directly in front of the listener. A sound source with 90° azimuth and 0° elevation lies directly on the right side of the listener. A sound source with 0° azimuth and 90° elevation lies directly above the listener.

Human listeners perceive the direction of a sound source the most accurately when the source is placed directly in front of them, that is, when azimuth is 0°. Errors in judgment of direction vary from 2° to 3.5° for sources in front [13]. When the sound source is moved to the side of, or behind the listener, the direction is estimated wrong even by 20°. An error as high as 20° sounds big and it seems that human auditory system does not perform as accurately at localizing objects as vision. However, before making hasty comparisons between auditory and visual systems, one should remember that vision covers only half (ca. 180°) of the full 360° circle that auditory system is able to cover, and the biggest errors in auditory localization occur for rear locations that are not even accessible for vision. Thus, the auditory system covers actually a wider area than vision, even if not accurately for all locations in space. In addition, one should keep in mind the physical properties of the ear and how they differ from those of the eye.

The difference between ears and eyes in the accuracy of localizing objects originates from the physical properties of these two organs. The image of the visual world is focused onto the surface of the retina by the optics of the eye. The neurons connected to the retina have an instant access to the spatial information about the objects that we see. An analogous way of processing the locations of sounds cannot be found in the ear because the neuronal representation of hearing begins with neurons in the ear that are specialized in processing frequencies of the sound rather than in extracting information about the location of sound sources [13]. Thus, the processing of the direction of a sound source has to be done in the brain by using completely different approaches compared to those of vision. The auditory system overcomes these challenges by making use of *auditory cues*, differences between the signals reaching each ear.

2.2 Auditory Cues for Localization

Consider a sound approaching the listener in a free sound field, that is, when the listener is not surrounded by walls or other hard objects that might reflect or diffract sounds. Depending on the direction of the sound source the approaching sound has to interact more or less with obstacles such as the listener's head, pinnae, and shoulders before reaching the eardrums. If the sound source is off to the side of the listener, the sound will also have to travel longer to reach the farther ear. Each obstacle modifies the frequency spectrum of an audio signal in a unique way and, thus, the sound detected by the left ear and the sound detected by the right ear are not identical. In addition, the distance delays the sound traveling to the farther ear more compared to the nearer ear, making the sound reach the ears asynchronously. These spectral and interaural differences are called *auditory cues*, and they are used by the auditory system in determining the directions of sound sources.

Sound localization relies on *binaural* and *monaural* cues. Binaural cues involve two ears, while monaural cues need only one ear. Binaural localization is based on the comparison of an audio signal detected by the two ears, and it is the most reliable way of localizing sound sources. To some extent it is also possible to localize sounds with one ear only. [11]

2.2.1 Interaural Time Difference and Interaural Level Difference

Interaural time difference (ITD) is the difference in the arrival time of a sound at each ear [11, 13]. Consider a sound coming from a sound source that is placed directly in front of the listener (Figure 2a). Since the sound source is at equal distance from both ears, the sound reaches the ears simultaneously. There is no time difference between the two ears and, thus, the ITD is 0. Now, assume that the sound source is moved closer to the right ear than to the left ear (Figure 2b). Since the path from the source to the left ear is longer than the path to the right ear, the sound reaches the left ear later resulting in a time difference between the two ears.

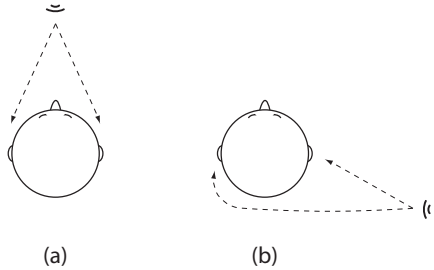

Figure 2: (a) The sound coming from the sound source in front of the listener reaches both ears simultaneously, and there is no time difference between the two ears (ITD = 0). (b) The sound coming from the sound source on the right side of the listener reaches the right ear before reaching the left ear. The sound signal reaching the left ear is delayed in time, resulting in an interaural time difference.

The ITD ranges from 0 to 690 µs depending on the location of the sound source. As stated above, the ITD is 0 for sound sources that lie directly in front of the listener. In fact, the ITD is 0 at any point on the median plane, because points on that plane are always at equal distance from the ears. When the sound source is moved more to the sides of the listener, the ITD increases, reaching its maximum value directly on the left and on the right side of the listener (azimuth is $-90°$ or $90°$ and elevation is $0°$). The value of the ITD decreases again when the source is moved from the side toward the back of the listener. Since the ITD varies as a function of sound source position, it can be used as an auditory cue to determine the direction of a sound source. [11]

The other binaural cue, the *interaural level difference* (ILD), is the difference in sound pressure level of a sound reaching the two ears. When not specified in decibels, it is referred to as *interaural intensity difference* (IID). ILD is not constant for all frequencies. Low-frequency sounds stay mostly unaffected by the listener's head because their wavelengths are longer than the dimensions of the head, as shown in Figure 3a. This is due to the *diffraction* of the sound which allows the sound waves to 'bend' around the listener's head. Frequencies that are low enough to pass the head unaffectedly are approximately equally intense at each ear. [11, 13]

However, the head casts an acoustic shadow for high-frequency sounds that have equal or shorter wavelengths compared to the dimensions of the head, preventing the sound waves from reaching the shadowed area (Figure 3b). High-frequency sounds do not get as close to the ear on the shadowed side as they do on the other side of the head. Higher frequencies attenuate at the shadowed ear and are, therefore, perceived less intense than on the opposite side of the head. [11, 13]

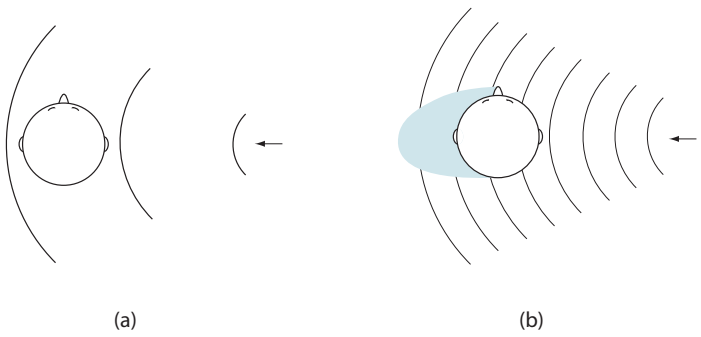

Figure 3: (a) Low-frequency sounds pass the listener's head unaffectedly. (b) High-frequency sounds are blocked by the listener's head, resulting in an acoustic shadow. In the shadowed area high frequencies are less intense compared to those on the right side of the head. The arrow indicates the direction of an incoming sound.

Like ITD, ILD is 0 when the sound source is directly in front of, behind, or above (or, less likely, directly below) the listener. This has to do with the nearly symmetric shape of the head and pinnae; when the paths from the sound source to the eardrums of the listener are nearly symmetric, differences in the levels of the

incoming sounds are inevitably nonexistent. For low-frequency sounds (< 500 Hz) the ILD is negligible, but it can be as large as 20 dB for higher frequencies. [11]

The theory behind ITD and ILD assumes that the head is a sphere with no external ears or asymmetry. Thus, there are several points in the coordinate system around the listener's head where both ITD and ILD are identical and cannot, therefore, be used to determine the direction of a sound source. When these points are mapped into the coordinate system, they form a cone that has its apex at the ear of the listener (Figure 4). This cone is called the *cone of confusion*. All sounds placed on the surface of this cone have the same ITD and ILD, so their locations cannot be distinguished. [11]

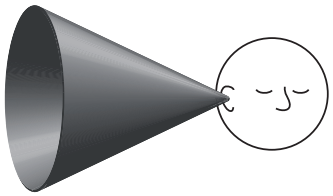

Figure 4: Cone of confusion. ITD and ILD are identical at all points on the surface of the cone. The locations of the sound sources placed on this surface cannot, therefore, be distinguished based on the interaural difference cues.

In summary, ITD is an important binaural cue in localizing sounds at low frequencies, while ILD provides more accurate information at high frequencies. The idea that the auditory system uses ITD at low frequencies and ILD at high frequencies to determine the direction of a sound source is often referred to as *duplex theory*, proposed by Lord Rayleigh in the beginning of the 20^{th} century. It does not, however, explain how the auditory system discriminates the front-versus-back locations, the directions on the surface of the cone of confusion, or the elevation of the sound. [14, 11]

2.2.2 Spectral Cues

The two binaural cues, ITD and ILD, are useful for the auditory system especially in determining the azimuth of a sound source. However, they are not enough for determining the elevation of a sound source. For instance, for sounds located in front of or behind the listener (0° or 180° azimuth), both ITD and ILD are zero regardless of the elevation angle. This happens because the paths from the sound source to each ear are equally long when the sound source is at any point on the median plane (Figure 1, dashed blue line). To obtain more accurate information about the elevation of a sound source, the auditory system relies mostly on *spectral cues*. Spectral cues also help to resolve the direction of the sound source on the surface of the cone of confusion.

Spectral cues are changes in the spectrum of an incoming sound, created by diffraction and reflection of a sound by the listener's head, pinnae and shoulders.

The interaction with the head, pinnae and shoulders causes the intensity of some frequencies to decrease and others to enhance. High frequencies (especially > 6 kHz) that have short wavelengths relative to the size of the listener's head are influenced by the head and especially by the small folds of the pinnae. Low frequencies are affected mostly by the shoulders of the listener. [11, 13]

The spectrum of a sound at the eardrums of the listener is unique for all directions of sound sources, because every combination of azimuth and elevation makes the sound reflect and diffract from the pinnae in a different way. Since the spectrum changes systematically as a function of the location of a sound source, spectral changes, or spectral cues, can be used by the auditory system in determining the direction of a sound source also when ITD and ILD yield unambiguous cues. [11]

2.3 Head-Related Transfer Function

The combination of the effects of ITD, ILD, and spectral cues on a sound coming from any given angle can be modeled with the *head-related transfer function* (HRTF). It is the Fourier transform of the *head-related impulse response* (HRIR) from the sound source to the eardrum, and it describes how the sound at the listener's eardrums has changed compared to the original sound. Since the size and the shape of the head and pinnae vary between individuals, the HRTF for a specific angle is slightly different for every listener. [11]

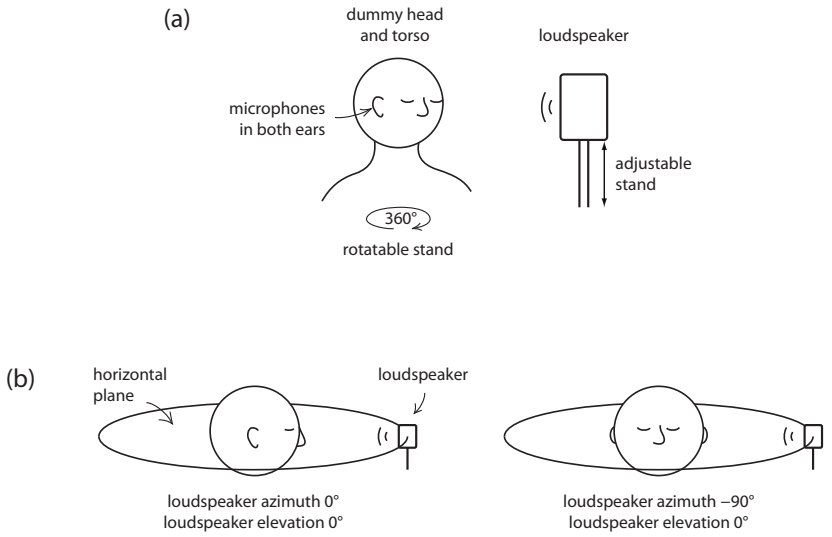

Figure 5: (a) HRTFs can be measured in an anechoic chamber using a dummy head and torso placed on a rotatable stand. Miniature microphones are placed at the ear canals of the dummy head. A full range of HRTFs for any elevation can now be recorded by playing sounds from the loudspeaker (adjusted to the desired height) and by rotating the dummy head in desired increments until the full circle (360°) of HRTFs has been measured. (b) An example of measuring the HRTFs at elevation 0° for azimuth 0° (left) and azimuth $-90°$ (right).

HRTFs can be measured in an anechoic chamber with small microphones placed at the entrances of the listener's ear canals. Sounds are then played from loudspeakers located at different angles around the listener's head. The recordings result in HRTF *pairs* (for left and right ears) for each recorded angle. Even though there is variation in the shape and the size of the pinnae between individuals, HRTFs are often measured with a *dummy head*, an artificial head and torso with pinnae (Figure 5). Dummy head measurements yield fairly good HRTFs that can be used to model the HRTFs of most listeners without having to measure them individually for everybody. [15]

2.3.1 Spatial Sound Synthesis

For a sound presented over headphones, the interaural differences (ITD and ILD) can be controlled accurately by simply adjusting the sound level and the delay between the left and right channels (the two ears). However, the sound usually lacks information about the shape of the listener's head, pinnae and shoulders, because it originates from the headphones and does not interact with the head, pinnae or shoulders. Without these spectral cues needed for localizing the direction of a sound source properly, the sound is perceived as having its origin within the listener's head rather than outside the head, typically somewhere along an imaginary line connecting the two ears. This phenomenon is called *lateralization*, in contrast to the term *localization* which refers to the more natural situation where the sound can be perceived as coming from any direction. [11]

Since HRTFs contain all the information needed for proper sound localization, they can be used to synthesize realistic 3D sound fields with spatial sounds that have a direction [16]. Any monaural source sound can be converted into a spatial sound by convolving it with the left and right HRTFs of the desired angle (Figure 6). The convolved pair is then presented binaurally to the listener over headphones, resulting in virtual spatial sound.

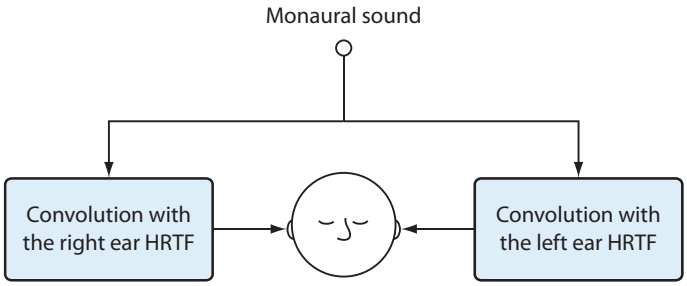

Figure 6: The principle of synthesizing spatial sounds. The original monaural source sound is convolved with the head-related transfer function (HRTF) pair of the desired angle separately for the left and right ears. The resulting sounds are presented binaurally to the listener over headphones.

Synthesized spatial sounds have some limitations. If the source sound is convolved with HRTFs that were obtained from dummy head recordings or with HRTFs designed for another person, the resulting 3D sound field may not be completely idealistic for a particular listener. All directions may not be recognized correctly. This is due to the differences between the shape of the listener's head and the head the HRTFs were designed for. However, the ability to localize sound sources correctly in the horizontal plane (azimuthal angles) does not deteriorate significantly even if non-individualized HRTFs, recorded either with a dummy head or with another person's head, were used [17]. Another problem related to the use of synthesized spatial sounds arises when the listener moves their head. Head movements distort the illusion of the 3D sound field, unless this is taken care of by constantly tracking the head movements [18].

It is also possible to synthesize a 3D sound field with a pair of loudspeakers. In this case a method called *crosstalk cancellation* must be used to prevent the unwanted signals from each loudspeaker from reaching the opposite ear. The problem in using loudspeakers is, however, that the auditory 3D illusion can be perceived only in a certain spot, called the *sweet spot*. Outside the sweet spot the illusion is lost. [18]

3 Functional Brain Imaging Techniques

Studying the functionality of the brain, the organ with billions of neurons, is a challenging task. *Single-unit recording* that gives the most accurate information about the firing of an individual neuron is a problematic recording method, because it is invasive and requires surgical operations. It is not painful, because the brain has no pain receptors, but it is potentially harmful to the brain. Therefore, it can only be applied on experimental animals, or on humans during brain surgeries that require single-unit recording due to medical reasons. Moreover, one cannot be sure how well the activity of a single neuron represents the activity of larger populations of neurons.

Several alternative, non-invasive brain imaging techniques have been developed to make it possible to observe the living brain without surgeries, and to reveal the activity in bigger neuronal populations than what single-unit recordings would allow. Some of these imaging techniques are *direct* measures of the brain activity, while others are *indirect*. Single-unit recording is a direct measure of the activity of a neuron, as the electrode measures the firing of an individual neuron. [19]

Electroencephalography (EEG) was the first non-invasive imaging technique. Like single-unit recording, EEG is a direct measure of neuronal activity. Unlike in single-unit recording, EEG measures the electric fields generated by the neuronal activity in the brain with electrodes that are placed on the scalp rather than inside the brain. The temporal resolution of EEG is excellent, because the electrical activity can be observed almost instantly. However, the spatial resolution is quite limited, because it is difficult to precisely locate the active neurons. In addition, it is not possible to measure the brain activity deep in the brain with EEG, because the electrodes will only detect the activity close to the surface of the scalp. [19]

Indirect brain imaging techniques such as *functional magnetic resonance imaging* (fMRI) and *positron emission tomography* (PET) provide better spatial resolution than EEG, but their temporal resolution is limited. This is because these techniques measure the brain activity indirectly, revealing the active areas by measuring the changes in the cerebral blood flow and blood oxygenation. This is based on the fact that the blood flow and the consumption of oxygen of active neurons increase. However, these changes are slow, occurring after a delay, which is why the temporal resolution of fMRI and PET is not good. [19]

None of the above recording methods can provide both temporally and spatially high resolutions at the same time. Sometimes different recording methods are used together to obtain the best aspects of certain techniques. There is, however, one method that has excellent temporal resolution and better spatial resolution than the EEG. This method is called *magnetoencephalography*.

3.1 Magnetoencephalography (MEG)

Magnetoencephalography (MEG) is a non-invasive brain imaging technique that measures the weak magnetic fields produced by the electrical activity of the neurons in the brain. MEG is the magnetic counterpart of the EEG, representing the magnetic fields produced by the brain rather than electric potentials. Like electric potentials in EEG, magnetic fields in MEG reflect the activity in the brain directly. [19, 20]

The temporal resolution of MEG is excellent (milliseconds). Similarly to EEG, MEG cannot measure areas deep in the brain and, thus, cannot compete with the spatial resolution of fMRI or PET, but its spatial resolution comes down to a few millimeters, which makes it spatially more accurate than EEG. This is because compared to electric fields, magnetic fields are less distorted by the resistance of the skull and the tissues.

In MEG, the brain activity is recorded from outside the skull with a helmet-shaped sensor array (Figure 7). The sensor array consists of *superconducting quantum interference devices* (SQUID) that are used to detect the extremely weak magnetic fields produced by a group of neurons. The magnetic fields are generated as a result of the changing electric currents in the neurons, much like the electric current in a wire induces a magnetic field around the wire.

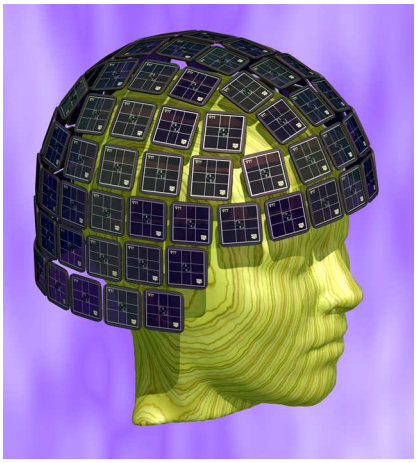

Figure 7: The helmet-shaped MEG measuring unit consists of dozens of superconducting quantum interference devices (SQUID) that are able to detect the weak magnetic fields of the brain. Each SQUID is responsible for recording multiple MEG channels. Image courtesy of Elekta.

Magnetic fields generated by the brain are considerably weak, only a few picoTeslas (1 pT = 10^{-12} T), or, as more commonly expressed, a few femtoTeslas per centimeter (1 pT = 1000 fT ≙ 10 fT/cm). The SQUID sensors are able to detect these faint magnetic fields, but they will also pick up noise from the surrounding environment. That is why MEG is typically recorded in a magnetically shielded room, which eliminates most of the magnetic fields of the environment. For example, the magnetic field of the Earth alone is 10^8 times stronger (microTeslas, 1 μT = 10^{-6} T) than the magnetic fields produced by the brain. However, the magnetic field of a single neuron is too weak to be detected even with the extremely sensitive SQUIDs and proper shielding. Therefore, each SQUID sensor within the MEG helmet measures actually a detectable combined magnetic field from a group of approximately 50 000 neurons. [19]

Despite the advantages, MEG also has some disadvantages compared to EEG. First, since the electrical activity of a neuron generates a magnetic field perpendicularly rather than in all directions around the neuron, only the neurons that are parallel to the skull produce a magnetic field that can be observed outside the head. Thus, MEG is able to detect only the magnetic fields generated by neurons (such as pyramidal neurons) that are in the correct orientation in relation to the skull, whereas EEG detects the electric currents of all neurons located near the scalp [19]. Second, the measuring equipment including the magnetically shielded room is expensive and cannot be moved around. That is one of the reasons why MEG is not as widely available as EEG.

3.2 Event-Related Fields

Event-related field (ERF) is a brain response, a faint magnetic field recorded from the scalp as a function of time, that is typically evoked by sensory, motor, or cognitive events. It is the magnetic counterpart of the *event-related potential* (ERP), the method used with the EEG. ERFs can provide valuable information about the functionality of the brain, because the magnitude of the evoked response is linked to the level of activity of the brain. Thus, ERFs allow us to study which brain areas are affected by different stimuli and what the level and the timing of this evoked brain activity is. [21, 22]

ERFs are relatively small in amplitude compared to the MEG signal, so they cannot be observed from an ongoing MEG signal [21]. As explained in the previous section, magnetic fields of the brain are extremely weak. If we present a stimulus to a subject and observe the continuous MEG signal measured by one SQUID sensor, we will see a graph that represents the current magnetic activity of the brain area covered by that sensor (Figure 8). It is hard to detect an ERF in the signal, because the neurons fire even when no stimuli are present. In addition, the SQUIDs are extremely sensitive and they pick up noise, no matter how properly the recording room was shielded from magnetic fluctuations of the environment.

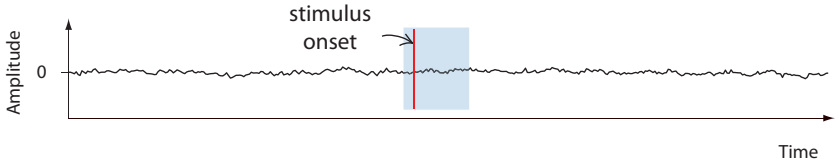

Figure 8: The helmet-shaped sensor array of the MEG device can consist of dozens of superconducting quantum interference devices (SQUID), each of them being responsible for recording multiple MEG channels. Here, a single MEG channel is shown. Due to noise and the background activity of the brain, no significant change can be seen in the amplitude of the magnetic field after presenting the stimulus (shaded area).

To make the small ERF responses visible, the activity of the brain is usually measured in relation to some event, for example, a sound presented to the subject. The same sound stimulus is typically repeated 100 times or more, and each time only the relevant time window, the section of the signal (between, for instance, 100 ms before and 500 ms after the stimulus onset) is analyzed (Figure 9a). Brain responses to the stimulus that are within the selected time window are then averaged with each other. Averaging makes the responses related to the stimulus emerge from the background activity of the brain and also cancel out the noise, leaving the ERF (Figure 9b). [21, 22]

ERFs that are evoked by auditory stimuli (*auditory evoked fields, AEF*) have typically a large component called the N1m (sometimes called the N100m according to its typical latency), which occurs approximately 100 ms after the stimulus presentation (Figure 9b). For the ERP, an equivalent peak is called the N1 or N100. N1m is usually the largest deviation in the amplitude of the response evoked by an auditory stimulus and it is believed to reflect early sensory processes and the allocation of neuronal resources. The amplitude of the N1m response can be modulated by various factors (physical parameters of the stimuli). For example, the intensity, the location, and the presentation rate of the stimulus can affect the magnitude of the amplitude [23, 2, 24]. Loud sounds, for example, elicit stronger N1m peaks compared to quieter sounds. Also selective attention is known to modulate the amplitude of the N1m [4].

The ERF has also other components, such as P1m and P2m, but they are less frequently used in studies involving the MEG. Peaks that occur after N1m (ca. 100 ms) but prior to 200 ms have been associated with late sensory and early perceptual processes, while cognitive processes such as language and memory are thought to be related to the peaks after 200 ms. The terminology of the ERF is derived from ERPs in EEG studies, where P1 and P2 denote the first and the second positive peaks of the ERP and N1 respectively the first negative peak. For the EEG the N1 peak is always negative, but for MEG it can be either negative or positive. [21, 22]

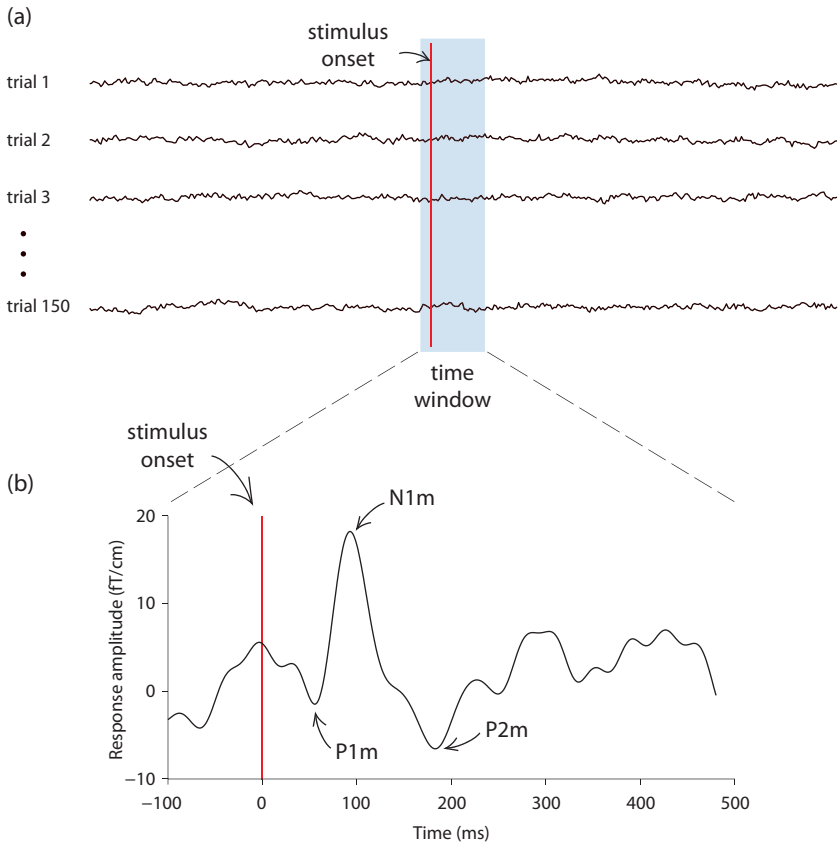

Figure 9: (a) The same stimulus (e.g., a sound) is presented in several trials (up to 150 times) and the relevant time window is analyzed for each trial. (b) An event-related field (ERF) is obtained by averaging brain responses to the stimulus over 150 trials within the stimulus-locked time window. The largest deviation occurring approximately at 100 ms after stimulus onset is called the N1m.

4 The Neuronal Representation of Auditory Space

As explained in Chapter 2, the human auditory system uses auditory cues such as ITD, ILD and spectral cues to determine the direction of a sound source. The neuronal representation of hearing begins with sensory neurons in the ear that are specialized in processing frequencies of the sound signal. The auditory nerve carries the signal from the inner ear to subcortical auditory nuclei and finally to the primary auditory cortex, where the frequencies are represented tonotopically by neighboring groups of neurons, much like the frequencies are represented in the basilar membrane of the cochlea of the inner ear [13].

In vision and touch the spatial receptive fields of sensory neurons are mapped in the primary cortices (primary visual cortex and primary somatosensory cortex) topographically in the same manner as frequencies are represented in primary auditory cortex. For example, if a certain area of the somatosensory cortex is stimulated, the person can feel their hand being touched. However, the role of the auditory cortex in extracting information about the locations of sound sources has been unclear, and it seems that there are not specific areas in the auditory cortex for specific sound locations. This is fundamentally different from other senses.

This chapter focuses on explaining how the auditory cortex participates in spatial processing of auditory stimuli, how the neurons sensitive to auditory cues are organized in the cortex, and how spatial location is encoded in the neuronal activity. In addition, the challenges and solutions for studying spatial sensitivity are discussed.

4.1 The Rate Code for Sound Azimuth in Monkey Auditory Cortex

Primates, such as monkeys, are used in brain experiments to get an insight into the functioning of the human brain. In a previous experiment, single-unit recordings were conducted in monkeys to study whether the spatial information carried by a sound signal is represented in the primary auditory cortex of monkeys [1]. Moreover, responses of individual neurons to different azimuths were studied.

The spatial sensitivity of several left and right primary auditory cortical neurons of three monkeys was measured while sounds with different azimuths were played from loudspeakers located around the monkey's head. The firing rates of individual neurons occurring during the stimulus presentation were used as a metric of neural activity. It was found that the spatial information carried by a sound signal is represented in primary auditory cortex of monkeys, but not in the same way as the spatial information of vision and touch are represented in the respective cortices.

Two different hypotheses for the neural coding of sound location in monkey auditory cortex were proposed. First, sound locations could be encoded in accordance with the *place code hypothesis*. In the place code hypothesis the neurons are assumed to be tuned to narrow ranges of, e.g., azimuths. This means that each neuron responds only to the auditory stimuli coming from a direction that falls in the particular neuron's narrow receptive field, i.e., within the range of sound az-

imuths that the particular neuron is specialized in recognizing. This would resemble the way vision and touch are represented in the respective primary cortices, with the only exception that the neurons in primary auditory cortex are assumed to be intermingled, rather than clustered together with similarly tuned neurons.

The second hypothesis proposed that sound locations could be encoded in accordance with the *rate code hypothesis*. The rate code hypothesis assumes that each spatially sensitive cortical neuron is tuned to one of two possible wide ranges of azimuths: the left or the right hemifield. In addition, each neuron codes the azimuth by firing at the fastest rate for auditory stimuli presented in the preferred hemifield and gradually slower as the sound source is moved farther to the non-preferred hemifield (i.e., lateral locations in the non-preferred hemifield). Similarly to the place code hypothesis, the rate code hypothesis assumes that the populations of neurons with distinct tuning properties (left and right hemifields) are intermingled.

The study indicated that the spatial sensitivity of auditory cortical neurons most closely resembled the rate code. All studied neurons within one hemisphere responded to nearly every sound location (all locations in both left and right hemifields), but the majority were more tuned to contralateral sound locations (sounds presented in the opposite hemifield) than ipsilateral sound locations (sounds presented in the same hemifield). This was seen also in the overall level of activity (averaged firing rate) of all recorded neurons within each hemisphere: the level of activity was higher for contralateral locations than for ipsilateral locations. However, the difference in the level of activity was not big (8–15%). It was also found that the neurons were indeed intermingled and not clustered together like spatially sensitive neurons are for other senses. Similar results have been found also in other experiments investigating the monkey auditory cortex [25, 26].

4.2 Challenges in Studying Spatial Sensitivity in Human Auditory Cortex

Studying the neuronal representation of sound source location in the human brain in a single-neuron level is challenging, because the most accurate results would require single-unit recording which cannot be applied to humans (see Chapter 3). The currently available non-invasive imaging techniques such as EEG, MEG, fMRI and PET measure the combined activity of larger groups of neurons. The spatial resolution of these imaging methods is not sufficient to detect the activity of individual neurons.

Problems arise especially when similarly responding neurons are not organized in clusters in the brain. First, the combined activation may not be straightforward to interpret if the proportions and the levels of activation of the underlying two populations of neurons are not known. Second, like in the study in monkeys, the difference in the combined level of activation may be so small that it might not be detectable with non-invasive imaging methods such as EEG and MEG [1]. This was suggested to be the reason for why previous studies in humans have failed to detect

any difference in the amplitude of evoked potentials as a function of sound source location. One solution to study the spatial sensitivity in human auditory cortex is to use the *stimulus-specific adaptation paradigm*.

4.3 Stimulus-Specific Adaptation Paradigm

The *stimulus-specific adaptation paradigm* provides a framework for measuring the brain responses (ERF and ERP) of two spatially overlapping populations of neurons with non-invasive brain imaging methods such as MEG and EEG that measure the combined activity of a group of neurons. In stimulus-specific adaptation neurons are thought to habituate to identical, repeated stimuli, resulting in smaller brain responses to the repeated stimulus, whereas novelty stimuli elicit larger brain responses. This paradigm has been used successfully in previous studies to measure the effects of frequency, intensity and sound source location on the ERP and ERF [27, 3, 2]. The paradigm is especially useful for studying the spatially sensitive auditory cortical neurons, because, unlike in vision and touch, auditory neurons with similar spatial receptive fields are not clustered together in the cortex [1, 2].

In a previous experiment, Butler used the stimulus-specific adaptation paradigm to study with EEG how varying the sound stimulus frequency and intensity affect the human vertex potential [27]. Two sound stimuli, a *test stimulus* and an *intervening stimulus*, were used in his study. He discovered that when a sequence of test stimuli was presented with occasionally occurring intervening stimuli, both having similar physical characteristics (constant frequency and intensity), the N1 responses to the test stimulus attenuated. In contrast, when either the frequency or the intensity of the intervening stimulus was changed, the N1 responses to the test stimulus increased. Butler concluded that stimuli with similar physical characteristics activated the same neural units. Thus, by sequentially presenting identical test and intervening stimuli, the activated neural units habituated to the sounds, which could be observed in attenuated N1 responses to the test stimulus. On the contrary, sound stimuli with different properties activated different neural units. Therefore, no habituation occurred and the N1 responses to the test stimulus remained strong.

In another experiment, Butler used the stimulus-specific adaptation paradigm to study how the location of a sound source affected the N1 responses of the test stimulus. He presented two sounds, a standard and a variable stimulus, sequentially to the subject. The standard stimulus was always in the same location, while the location of the variable stimulus was altered. The results were similar to his previous studies: when the two auditory stimuli were near one another, the response amplitudes were smaller than those of the stimuli that were widely separated in space. [3]

The basic interpretation of the stimulus-specific adaptation paradigm for sound source location is illustrated in Figure 10. Assume that there are two equally large populations of spatially selective auditory cortical neurons in each hemisphere. One population responds to sounds presented in the left hemifield (red), while the other one responds to sounds presented in the right hemifield (blue). When two identical sounds are presented sequentially in the same location (left–left), they activate the

same neurons (Figure 10a, middle). The neurons adapt to the first sound by decreasing their level of activity, so when the sound is presented for the second time, the neurons do not respond as actively as to the first sound. With MEG this is perceived as an attenuated N1m response to the second sound (Figure 10a, bottom). Similar results are obtained when the two sounds are presented in the right side (right–right, Figure 10c).

In contrast, when the two sounds are presented in different locations (left–right), they activate different neurons (Figure 10b, middle). The N1m response to the second sound stays large, because the population activated by the second sound has not been activated recently and, thus, no adaptation has occurred (Figure 10b, bottom).

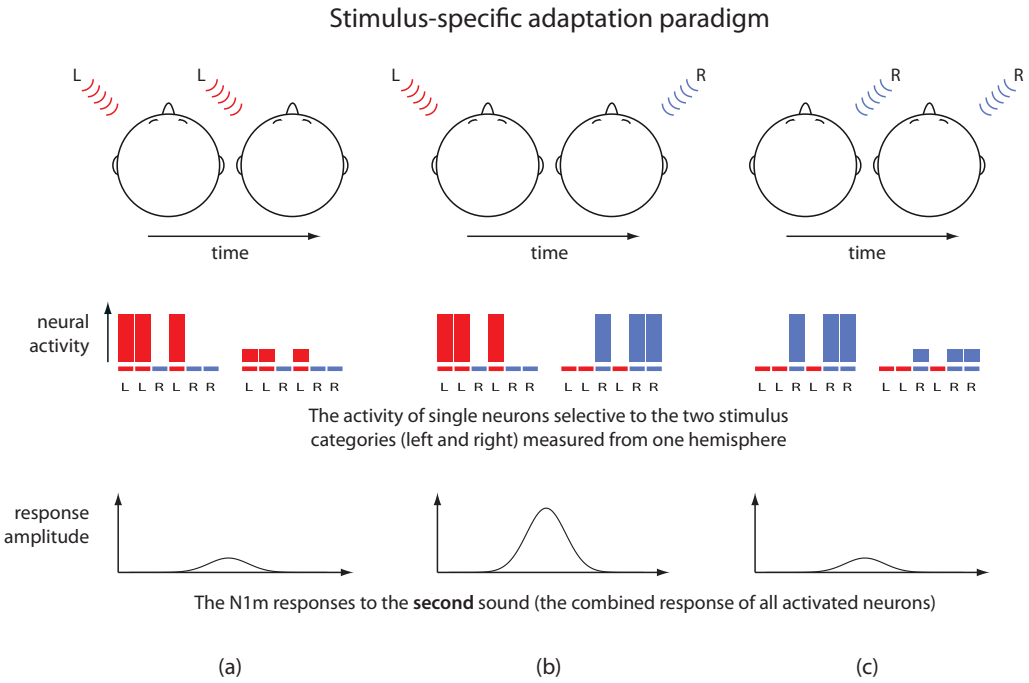

Figure 10: Stimulus-specific adaptation paradigm: (a) When two sounds are presented in the same location (left–left), they activate the same neurons (red). Due to the adaptation the neurons will not respond to the second sound presented in the same location as actively as to the first sound. This leads in maximal attenuation of the N1m response (bottom) measured for the second sound. (b) When the two sounds are presented in different locations (left–right), they activate different neurons (red and blue). The response is now larger, because the neurons that are activated by the sound coming from right have not been recently activated and, thus, no adaptation occurs. (c) Similarly to the left–left condition, the attenuation is maximal when the sounds are presented in the same location (right–right).

Finally, let us forget Figure 10b for a moment. We will notice that when the two sounds are presented in the same location (left–left or right–right), the N1m responses are equally large. This can happen even though different neurons are causing the activation, because both populations are of the same size, causing similar overall activations. Thus, by simply comparing these two N1m responses it is impossible to conclude whether there are spatially selective neurons in auditory cortex that would respond to the sound location as a function of azimuth. However, when the sounds are presented in different locations, the existence of these populations can be revealed with MEG (Figure 10b).

4.4 The Population Rate Code in Human Auditory Cortex

Similar results to the study in monkeys (see Section 4.1) were obtained in a recent experiment where the spatial sensitivity of human auditory cortex was measured non-invasively with MEG. The stimulus-specific adaptation paradigm (see Section 4.3) was successfully used to confirm that there are indeed two spatially selective populations of neurons in the human auditory cortex, one responding preferentially to sounds to the left and the other to the right of the listener. The majority of the neurons were found to be tuned to sounds presented in the contralateral hemifield. [2]

The procedure of the experiment was similar to the previous study conducted by Butler [3] with varying sound source locations. Two sounds, an *adaptor* and a *probe*, were presented sequentially to the subject from two locations, and the N1m response to the probe sound was measured. It was found that the N1m response varied systematically as a function of the spatial separation between the two stimuli. However, sounds originating from the same hemifield caused similar attenuation to the N1m response independent of the spatial separation. In contrast, the N1m response was the most prominent when the sounds were presented in different hemifields.

It was proposed that the *population rate code of auditory space* exists in human auditory cortex. That is, the sound locations are encoded in the relative level of activity of the two neuronal populations, one activated preferentially by sounds to the left of the listener and the other by the sounds to the right of the listener, as illustrated in Figure 11. The majority of the neurons within one hemisphere are tuned to contralateral sound locations and, thus, the effect of the ipsilateral preferring population to the overall activity is smaller.

The population rate code is consistent with the generally accepted fact that the discrimination of sound source locations is the most accurate for frontal locations [13]. When two sounds are presented in different locations in front of the listener where the curves are the steepest, even a small separation between the sounds makes a big difference in the level of activity of the neurons. In contrast, for two sounds with a similar separation presented more to the left or right side of the listener the difference in the activation of the neurons is not as large, so the spatial separation is more difficult to detect. The population rate code is, however, only suitable for explaining the coding of sound source locations in the horizontal plane [2].

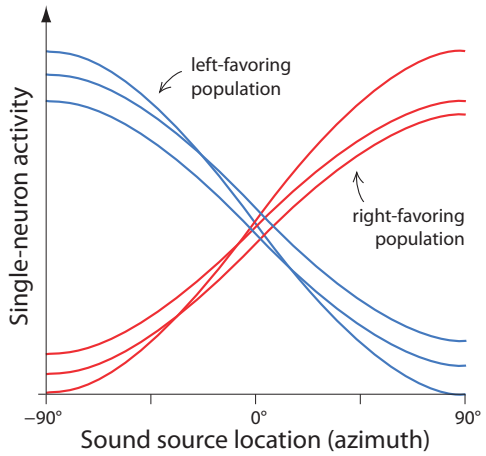

Figure 11: The population rate code. The human auditory cortex has two spatially sensitive populations of neurons: one is activated preferentially by sounds coming from the right hemifield (red) and the other by sounds coming from the left hemifield (blue; for simplicity, only three neurons from each population are presented). The majority of the neurons are tuned to the contralateral hemifield. Therefore, if these curves represented the neuronal populations in the right auditory cortex, most of the neurons would belong to the left-favoring blue population. The discriminability of the directions of two sound sources is the best straight in the front (at azimuth 0°), where the curves are the steepest. When two sounds are presented close to the midline, they fall at different heights of the curves. When the same sounds are presented more to the side of the listener, where the curve is more flat, they activate the main population of that hemifield nearly equally intensively, making it difficult to tell apart the directions.

5 Selective Attention and Auditory Cortical Processing

Selective attention is a cognitive process that enables one to process relevant inputs and ignore irrelevant ones. Perhaps the most commonly cited example of selective attention is the *cocktail party effect*, studied by the British psychologist E. C. Cherry in 1953. Cocktail party effect means the ability of selectively attending to one speaker in a crowd while ignoring the others, even though multiple competing and equally intense speech inputs reach the ears simultaneously. [22]

Attention can be divided into two main categories: top-down *voluntary attention* (endogenous or task-dependent process), and bottom-up *involuntary attention* (exogenous or stimulus-driven process). Attention is voluntary when it is directed to something under control of an individual and involuntary when it is attracted by a sudden external stimulus. In addition, the terms *overt* and *covert attention* are used. Overt attention means directing the sense organs toward the stimulus, for example, by turning the head (ears), or by shifting the gaze (moving the eyes) toward the location of interest. In contrast, covert attention is the act of mentally focusing on a relevant stimulus despite the position of the sense organs. [22]

5.1 The Effects of Unimodal Selective Attention

It is well known that when attention is directed within a single modality (i.e., unimodally), the neuronal activity in the corresponding modality-specific areas of the cortex increases compared to the unattended condition. For example, selectively attending to auditory stimuli increases the activity in the auditory cortex. In voluntary selective attention, the increased neuronal activity often correlates with improved behavioral performance, such as shortened reaction times and better stimulus discrimination of the attended stimuli. Such correlations have been found in both monkeys [28] and humans [29]. The increased neuronal activity is believed to reflect the enhanced neural processing reserved for the relevant stimuli.

5.1.1 Attention in Dichotic Listening

The effects of auditory selective attention on the neural activity in human auditory cortex have been studied extensively. In many ERP and ERF experiments attention was studied using the *oddball detection task* (e.g., [4, 5]). An auditory oddball detection task typically consists of randomly presented sounds of which the majority are standard stimuli (e.g., a 1000 Hz tone) and a fraction target stimuli (e.g., a 1050 Hz tone). Usually two competing auditory streams with unique standard and target stimuli are delivered dichotically over headphones to the listener's ears while only one ear is attended at a time. The electroencephalogram or the magnetoencephalogram is recorded while the subject is responding to the target stimuli whenever they appear at the attended ear, for example, by pressing a button.

In EEG and MEG studies, the attentional modulation of information processing is often investigated with the ERP (ERF for MEG) subtraction waves, such as the *negative difference* (Nd) and the *processing negativity* (PN) [30, 31, 32]. Nd and PN represent the difference of the ERP waveforms between attended and unattended conditions. Subtracting the ERP evoked by an unattended stimulus from the ERP evoked by an attended stimulus gives the Nd (or PN) differential (Figure 12). Nd and PN are often seen as indicators of attention-related enhancement in the neural processing. However, the effect of attention can also be measured as the amplitude of the N1 (N1m for MEG) peak itself.

EEG studies using variations of the oddball detection task have shown that the N1 responses peaking typically at 100 ms after stimulus onset are larger for attended auditory stimuli and smaller for unattended auditory stimuli (or the Nd/PN difference waves show a change between attended and unattended conditions) [4, 33, 34]. Similar results have been found in MEG studies for the N1m [5, 35]. Signs of increased activity in the auditory cortex have also been reported in studies using the fMRI [36, 37, 38] and the PET [39, 40].

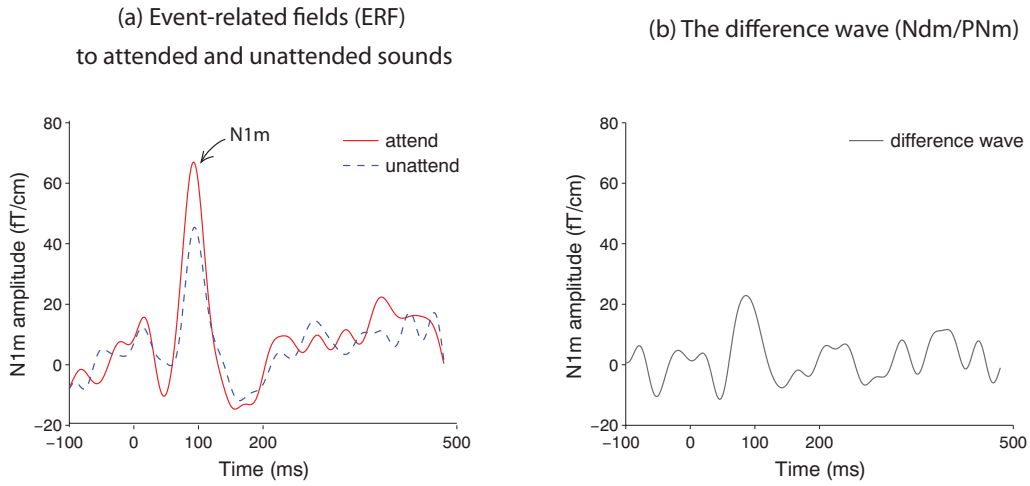

Figure 12: (a) Event-related fields (ERF) to an auditory stimulus recorded with MEG in attended and unattended conditions. These waveforms have also negative components, because they have not yet been summed with adjacent channels (see Figure 23). The first prominent peak is the N1m. (b) The magnetic counterpart of the negative difference, Ndm (also known as the processing negativity, PNm), is obtained by subtracting the ERF evoked by an unattended stimulus from the ERF evoked by an attended stimulus.

5.1.2 Auditory Spatial Selective Attention

As stated before, auditory selective attention can be directed covertly (i.e., mentally, without turning the head toward the stimulus) to a certain spatial location in auditory space. The above experiments, however, studied the influences of selective attention on auditory cortical activity using dichotic listening where independent streams of auditory stimuli were delivered to the listener's ears over headphones. These experiments did not use real spatial sounds as the stimuli but sounds that lack the auditory cues (see Section 2.2).

Other previous studies have, in contrast, used an array of loudspeakers in order to investigate spatial auditory attention with real spatial sounds [41, 42, 6]. These studies have provided evidence that sounds that are presented in an attended location evoke larger N1 amplitudes than sounds presented in unattended locations. These results were found by delivering tones or noise bursts (of which, e.g., 90% were standard, and 10% deviant stimuli) to the subject in random order via an array of loudspeakers placed at equal distance around the subject's head. In a typical experiment, the subject's task was to attend to one of the loudspeakers (e.g., leftmost, center-most, or rightmost) and press a button to the deviant stimuli whenever they were presented at the attended location. At the same time, event-related brain responses (ERP) were measured using the EEG from the subject's scalp. In all of these studies the main finding was that the amplitude of the N1 peak was larger for stimuli that occurred at attended locations than for those occurring at unattended locations. Thus, attending to a specific location in auditory space seems to enhance the processing of auditory stimuli occurring in the attended location.

5.2 Cross-Modal Attention between Audition and Vision

Studies on selective attention have traditionally focused on a single sensory modality at a time, although in most everyday situations it is rare that we would have to direct attention within just one modality, e.g., within just hearing or just vision. Indeed, we are typically surrounded by multiple stimuli competing for the attention of our sense organs at the same time. Let us revisit the cocktail party effect mentioned in the beginning of this chapter. In a crowded and noisy place one has to selectively attend to the most relevant stimuli in order to successfully follow a person speaking. In such situation, the relevant stimuli probably consist of both the lip movements (reading from the lips) and the speech signal originating from the speaker's mouth. In this example, cross-modal (i.e., involving more than only one sense) interactions between hearing and vision are needed to successfully combine the attended signals into a speaking and moving person.

As explained in Section 5.1, selectively attending to stimuli within a single modality strongly modulates the neural activity in the respective modality-specific cortical area. For example, attending to sounds increases the activity in the auditory cortex. However, there have been mixed results about what kind of cross-modal links exist between visual selective attention and auditory cortical activity.

5.2.1 Cross-Modal Auditory Spatial Selective Attention

Unimodal studies have proven that spatial selective attention exists within the auditory modality, but these studies did not shed light on whether cross-modal links in spatial attention between audition and vision exist. Links similar to the unimodal findings have been found between audition and vision in one of the first cross-modal ERP studies [8]. In that study, flashes and sound bursts were presented to the subject either in the left or the right side. The subject's task was to attend to one side, and one stimulus type (flashes or sound bursts) at a time. It was found that sounds that were presented at the attended location elicited enhanced negative deviations regardless of which modality was used to direct attention to that location. In contrast, the negative deviations were smaller for the sounds that were presented at unattended locations.

Similar results have been found in a more recent ERP study where cross-modal links in spatial attention between vision, audition, and touch were extensively investigated [7]. An enhanced negativity of the early negative difference (Nd), occurring before 200 ms, was found in the auditory-evoked potential (AEP), not only when attention was directed to a specific location within the auditory modality, but also when visual stimuli occurring at spatially congruent locations with the auditory stimuli were attended. Since the early Nd is thought to originate from the auditory cortex, it was suggested that attention-related changes in this negativity proved that visual spatial attention has an effect on auditory cortical processing. In addition, similar results were found for touch and audition, implying that cross-modal links in spatial attention exist also from touch to audition.

5.2.2 Cross-Modal Effects Between Synchronized Auditory and Visual Stimuli

Presenting visual stimuli synchronized with auditory stimuli can have unexpected effects. First, when a visual stimulus is presented synchronized with a spatially disparate auditory stimulus, it can have a strong biasing effect on one's ability to correctly localize the auditory stimulus. This is called the *ventriloquism effect* and it leads to mislocalization of the sound toward the location of a simultaneously presented visual stimulus [43, 44]. The ventriloquism effect was reported to bias the ability of correctly localizing the auditory stimuli even with disparities as great as 20° [44]. Moreover, it was found that the bias was the greatest when the visual stimuli were in the center of the visual space.

Second, synchronized visual and auditory stimuli tend to form multisensory objects. It has been proposed that attention can spread across modalities and even space if the attended modalities are a part of a multisensory object defined by synchronous but spatially disparate auditory and visual stimuli. In a previous study it was found that attending to visual stimuli enhanced the neuronal responses to the simultaneously presented auditory stimuli even though the auditory stimuli were task-irrelevant and presented in an unattended location [45]. It was concluded that when one modality of a multisensory object is attended, also other modalities related to this object receive enhanced processing. This result implicates that presenting

stimuli from one modality synchronized with spatially disparate stimuli from another modality is enough to form multisensory objects.

5.2.3 The Effect of Short-term Memory Load

In a previous MEG study it has been proposed that the level of short-term memory load, rather than visual attention, enhances brain responses to unattended auditory stimuli [46]. The subjects were instructed to memorize visually presented digits accompanied with irrelevant tones while the N1m response amplitudes to the irrelevant tones were measured with MEG. It was found that the N1m response amplitudes elicited by task-irrelevant auditory stimuli were larger during the visual memory task compared to a no-memory-task condition. In addition, it was found that the processing of auditory stimuli depended on the level of short-term memory load: easy memory tasks enhanced the auditory N1m response amplitudes less than difficult memory tasks. The enhancing effect of the visual short-term memory task was present even though the subjects were instructed to ignore the auditory stimuli. The results indicated that the short-term memory task presented in the visual modality increased the neuronal activity in the auditory cortex, enhancing brain responses to unattended auditory stimuli.

5.2.4 Findings of the Suppressing Effects of Attention

In some previous studies using fMRI and PET it was reported that attending to a visual task decreased the neuronal activity in the auditory cortex [10, 9]. The writers of the latter study concluded that attending selectively to visual stimuli may be linked to the decreased activity in areas that process stimuli from unattended modalities [9]. However, this claim is controversial, and less clear results about decreased activity in task-irrelevant cortical areas have been found in other studies, suggesting that the possible decrease in the task-irrelevant modality may be associated with other factors such as arousal rather than stimulus processing [47, 48]. It has also been noted that in studies where the suppression of the unattended modality was proposed, either no stimuli were presented in the unattended modality to assess the unattended modality, or the increased activity in the attended modality was not distinguished from persistent activity (i.e., information being processed in working memory) [48, 49].

6 The Aims and Hypotheses of the Study

The primary interest of this thesis was to investigate how attending to visual stimuli affects spatial processing in human auditory cortex. As the experiment involved human subjects, the invasive single-unit study was not an option in exploring the activity of the brain. Of the non-invasive brain imaging methods, the MEG was used, because its spatial resolution is better than that of the EEG, and its temporal resolution is excellent compared to that of other brain imaging methods, such as fMRI and PET.

In this study, brain responses to auditory stimuli were measured with and without the presence of visual selective attention. To draw the visual attention overtly to the appropriate direction (by moving the eyes either toward or away from the auditory stimuli), a visual 1-back task was used. To ensure that the level of attention was sufficient for the purposes of the experiment, the task was designed to be challenging enough. Thus, if necessary, the brain responses recorded from subjects who failed in the task could then be discarded due to the poor level of attention.

In a previous study it was found that when sounds and visual stimuli were presented synchronously in time but in disparate locations, the perceptual origin of the sounds was shifted toward the visual stimuli (ventriloquism effect) [43, 44]. In addition, another study reported that visual and auditory stimuli can form multisensory objects if presented synchronously [45]. This was found to lead to enhanced neuronal processing in the unattended modality even though the stimuli were spatially disparate compared to those of the attended modality. In the present study, auditory and visual stimuli were deliberately presented asynchronously to avoid the ventriloquism effect and the formation of multisensory objects caused by the synchronization.

The present experiment was designed to activate primarily neurons in the right auditory cortex, because previous studies have shown that the right auditory cortex is more sensitive to the spatial sound stimulation, leaving the responses measured from the left auditory cortex usually weaker [50, 51]. To activate the right-hemispheric neurons, most of the auditory stimuli were presented in the left side of the listener, because it has been previously shown that the majority of the spatially selective neurons within one hemisphere are activated preferably by contralateral sounds (sounds presented in the opposite hemifield) [1, 2]. Thus, the brain responses recorded from each hemisphere are expected to be stronger for contralateral sounds and weaker for ipsilateral sounds. Furthermore, the brain responses recorded from the left hemisphere are expected to be weaker than those recorded from the right hemisphere.

Stimulus-specific adaptation paradigm was chosen as the method to reveal the possible effect of visual attention on spatial processing in the auditory cortex. This paradigm has been used previously to reveal the otherwise invisible neural coding (population rate code) of the auditory space [2]. In this paradigm, two sounds, an adaptor and a probe, are presented in two different locations. The effect of the preceding adaptor on the response to the subsequent probe depends on the separation between these two sounds: the bigger the separation, the larger the brain

responses to the probe will be. Thus, this paradigm provides a frame of reference for the present study, making it possible to elicit auditory brain responses, affected by attention, in a spatial context. Without this or a similar paradigm it would be difficult to tell to what extent the possible variation in the amplitude of the N1m is caused by the two opponent spatially sensitive neuronal populations, and to what extent it is modulated by attention.

The probe will be presented in the left side of the subject throughout the experiment, while the location of the adaptor will vary (left–center–right). The brain responses to the adaptor are expected to be weaker in the right hemisphere than the responses to the probe, because only one of the possible adaptor locations (left) is clearly supposed to activate the majority of the contralaterally tuned neurons in the right hemisphere. Similarly, the responses to the probe are likely to be nonexistent in the left hemisphere but prominent in the right hemisphere, because the probe is always presented in the left hemifield and, thus, supposed to maximally activate the dominating left-tuned population in the right hemisphere.

The effect of the population rate code is expected to be seen clearly in the brain responses elicited by the probe when no attention to visual stimuli is present, as this will resemble the experimental set-up from a previous study [2]. When visual attention is present, it is possible that the pattern of the population rate code is modulated. Based on the knowledge from previous studies, such as [7, 8], it is likely that attending to the visual stimuli presented either in the same or disparate location with the auditory stimuli will increase the brain responses (the N1m) to spatially coincident auditory stimuli and, on the other hand, decrease the responses to spatially disparate auditory stimuli. Thus, since the probe is presented in the left, it is likely that the responses to the probe are larger when attention is directed to the left than when it is directed to the right.

7 Materials and Methods

This chapter will describe the present experiment that was conducted in order to find answers to the questions arisen in the previous chapter. The participants, the stimuli, and finally the task and procedure will be described in detail in the following sections.

7.1 Participants

Fourteen paid, healthy volunteers participated in the experiment. A written informed consent was obtained from each subject. All subjects reported having normal hearing and normal vision (or corrected to normal with contact lenses). Two subjects were excluded due to the poor signal-to-noise ratio (SNR), that is, the brain responses recorded from the cortex by the sensors of the MEG device were too weak. Twelve subjects of which seven were male were taken in the final analysis (ages 20–30 years, $\overline{x} = 24$, $\sigma = 2.8$). One of the subjects was left-handed.

7.2 Stimuli

Both auditory and visual stimuli were used in the experiment. The stimuli were presented to the subject by using the Presentation software (Neurobehavioral Systems, Inc.).

7.2.1 Auditory Stimuli

The auditory stimuli used in the experiment consisted of two alternating spatial sound stimuli, an *adaptor* and a *probe*, that were presented to the subject over earphones. The adaptor had three possible azimuths (−20°, 0° and 20°), while the probe was presented in a fixed location (azimuth −20°) as illustrated in Figure 13. The elevation was 0° for both the adaptor and the probe.

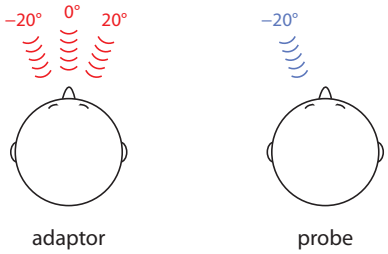

Figure 13: The possible azimuth angles for the adaptor (−20°, 0° or 20°) and the probe (−20°). The probe is identical with the adaptor presented in −20°. The elevation for both the adaptor and the probe was 0°.

All adaptor and probe stimuli were 200-ms sound bursts (44.1 kHz, 16 bit, stereo) of white Gaussian noise (including 20 ms linear rise and fall time), delivered to the subject over earphones. For the purposes of the experiment, it was seen as necessary to be able to present the auditory stimuli in the same virtual spatial location with the visual stimuli. Regular stereo sounds (sounds with ITD and ILD) lack the spectral cues and would, thus, be located within the head. Therefore, spatial sound stimuli were used. Another way to present sounds with the essential auditory cues is to use an array of loudspeakers placed at equal distance around the subject's head. However, this was not an option, because loudspeakers would have interfered with the MEG recording.

The spatial sound stimuli were generated in Matlab (The MathWorks, Inc.) by convolving white Gaussian noise with the HRTF of the desired angle. The HRTFs were obtained from the open HRTF database provided by Massachusetts Institute of Technology [52, 15]. The HRTFs in this database were recorded with the KEMAR Dummy-Head Microphone. The database contained two extensive sets of HRTFs for different elevation and azimuth combinations: the full data set recorded with both a small and a big ear, and the compact data set recorded with a small ear only. The compact data set was used in the present experiment, because the data in that set had already been equalized to compensate the non-uniform response of the loudspeaker used in the HRTF recording.

Special auditory tubes with tips resembling foam earplugs were used to present the auditory stimuli, because regular earphones would have interfered with the MEG recording (Figure 14). The auditory stimuli were set at a comfortable sound pressure level (well above threshold). To minimize the level difference that placing the earplugs asymmetrically might have caused between the left and right channels, the volume was adjusted individually for both channels up until the subject reported that both channels sounded equally loud.

Figure 14: Special MEG-compatible auditory tubes with disposable foam tips were used to deliver the auditory stimuli to the subject's ears.

The presentation of the sounds followed the principle of the stimulus-specific adaptation paradigm (see Section 4.3). The two auditory stimuli, the adaptor and the probe, were presented repeatedly in sequential adaptor-probe pairs with an onset-to-onset interstimulus interval (ISI) of 1 s. In total, three possible adaptor-probe pairs were obtained by combining the possible angles of the probe and the adaptor (Figure 15). These three pairs were exclusionary, that is, the pairs were never mixed during one trial.

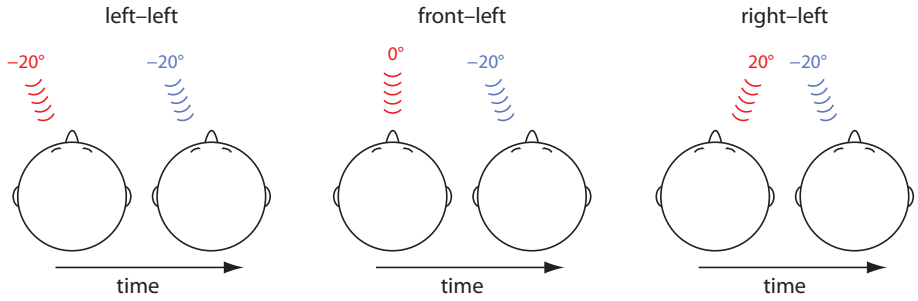

Figure 15: The three different adaptor-probe pairs used in the experiment: left–left, front–left and right–left. The adaptor (red) was presented either from the left, front, or the right side of the listener, while the probe (blue) was always presented from the left side. The ISI (onset-to-onset) of any two sounds was a constant 1 s.

7.2.2 Visual Stimuli

The visual stimuli were rear-projected through a hole in the wall of the magnetically shielded room onto a translucent white projection screen (93×70 cm) placed inside the magnetically shielded room. The visual stimuli consisted of six solid-color circles that were projected with a black background on the screen. The colors of the circles were chosen so that they would be easy to distinguish (Figure 16). Each circle was approximately 3.7° in diameter (visual angle).

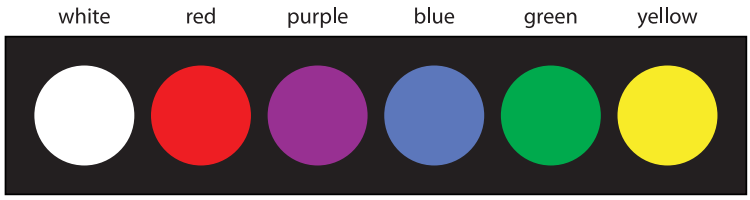

Figure 16: The visual stimuli consisted of six colored filled circles (white, red, purple, blue, green, and yellow). Each circle was approximately 3.7° in diameter (visual angle).

The circles were presented sequentially on the screen in a rapid pace with a constant onset-to-onset ISI of 650 ms. Each circle was visible on the screen for 150 ms. The order of the circles was randomized with one limitation: two consecutive circles were allowed to have the same color only with a 5% probability.

In order to direct the attention of the subject to a desirable direction, the circles were presented on the screen either on the left or the right side of the subject but never on both sides within the same trial (Figure 17). The circles were positioned in a 20° viewing angle on either side to align them with the outermost auditory stimuli in space. In addition to the circles, a fixation point (a white cross) was shown on the screen during the entire trial to mark the center of the screen.

Due to the zooming limitations of the projector, the screen was placed at 93 cm distance from the subject's eyes in order to obtain the desired 20° viewing angle. The circles were aligned vertically with the subject's ear level, which resulted in the circles being slightly below the horizontal midline of the screen.

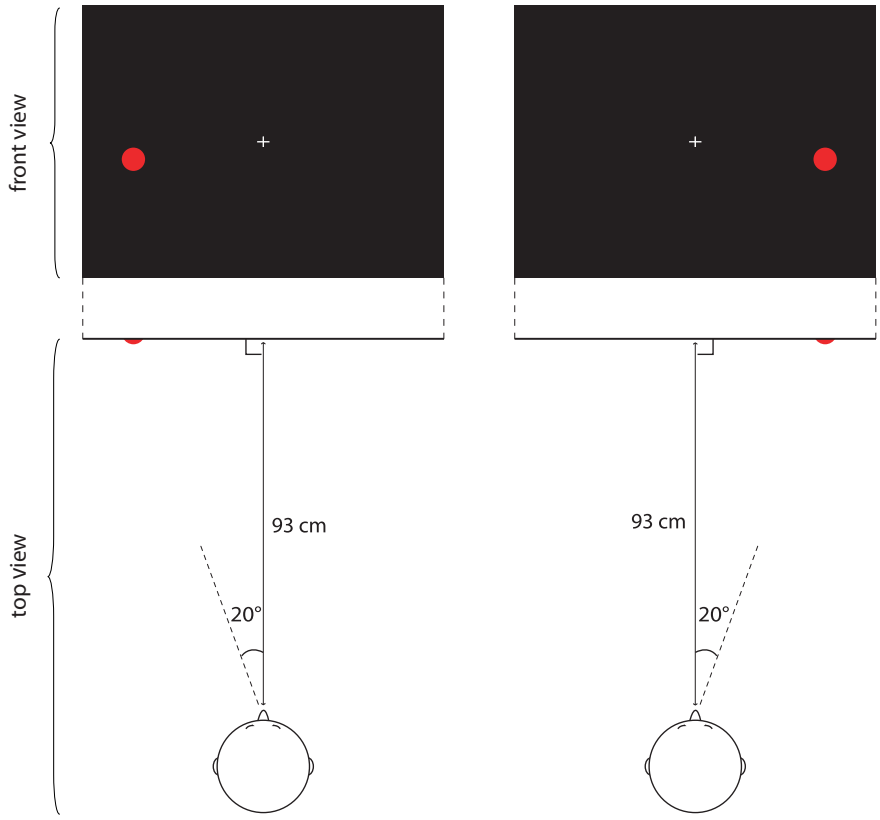

Figure 17: The visual stimuli (colored filled circles) were presented on a black background in a 20° viewing angle either on the left or on the right side of the subject. In addition, a fixation point (white cross) was shown on the screen during the entire trial to mark the center of the screen.

7.3 Task and Procedure

The experiment consisted of three conditions: attend to left, attend to right, and a passive condition, during which the subjects read silently a book of their choice. Due to the three possible adaptor-probe angle combinations (Figure 15) and the three different conditions, nine trials (nine different angle-condition combinations) were obtained (Table 1). Each trial lasted approximately five minutes.

Table 1: The nine trials of the experiment: the azimuth angles for the adaptor and the probe, and the condition. The trials were obtained by combining the three conditions (attend to left, attend to right, and passive) with the three different adaptor-probe pairs.

Trial	Adaptor	Probe	Condition
1	−20°	−20°	Attend to left
2	0°	−20°	Attend to left
3	20°	−20°	Attend to left
4	−20°	−20°	Attend to right
5	0°	−20°	Attend to right
6	20°	−20°	Attend to right
7	−20°	−20°	Passive
8	0°	−20°	Passive
9	20°	−20°	Passive

In trials where attention was required, colored circles were presented sequentially on the screen in 20° viewing angle in random order either on the left side of the subject (attend to left) or on the right side of the subject (attend to right). The subject was instructed to point their nose toward the fixation point located in the middle of the screen, but they were allowed to move their eyes toward the appropriate direction (overt attention). At the same time the two alternating sounds, the adaptor and the probe, were presented repeatedly to the subject over earphones, as illustrated in Figure 18.

The subject's task was to react to the target stimuli (two consecutive circles having the same color, a 1-back task) by lifting their left or right index finger briefly off the response pad. The subject was under the instruction to ignore the auditory stimuli. In trials where attention was not required (passive), the screen was turned off and only the auditory stimuli were delivered to the subject's ears over earphones. The subject was under the instruction to silently read a book or other material of their choice and ignore the auditory stimuli.

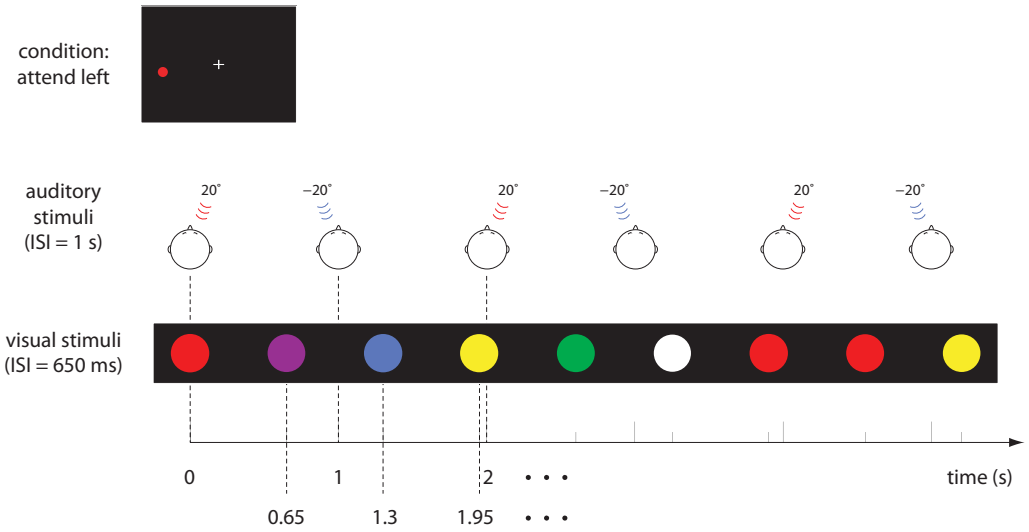

Figure 18: An example of a temporal sequence of the stimulus-presentation order in attend to left condition (Table 1, Trial 3). The auditory stimuli were presented in adaptor-probe pairs (right–left) with an onset-to-onset ISI of 1 s. The visual stimuli were randomized and presented with an onset-to-onset ISI of 650 ms on the left side of the subject. The subject was instructed to ignore the auditory stimuli and respond with the response pad to any two consecutive circles having the same color.

The order of the nine trials was carefully randomized for each subject and run successively with only the necessary breaks. However, the presentation order was locked to one of the two possible types illustrated in Figure 19. The first three trials and the last three trials always required attention either to the left or to the right, and the three trials in between always consisted of the trials with passive condition.

The passive trials were placed in the middle of the experiment, because more than three consecutive conditions requiring attention would have strained the eyes unnecessarily. Due to the limited time reserved for each measurement, the passive conditions were not placed for example after every two trial requiring attention, because the door to the magnetically shielded room would have had to be opened each time. This would have prolonged the experiment unnecessarily. Thus, the passive trials were placed in the middle of the experiment. In addition, two trials requiring attention to the same direction (e.g., attend to left – attend to left) never followed each other.

The experiment lasted approximately two hours including preparations. The MEG measurement itself lasted on average 65 minutes including short pauses during which the subject was allowed to fix their position or have a short break.

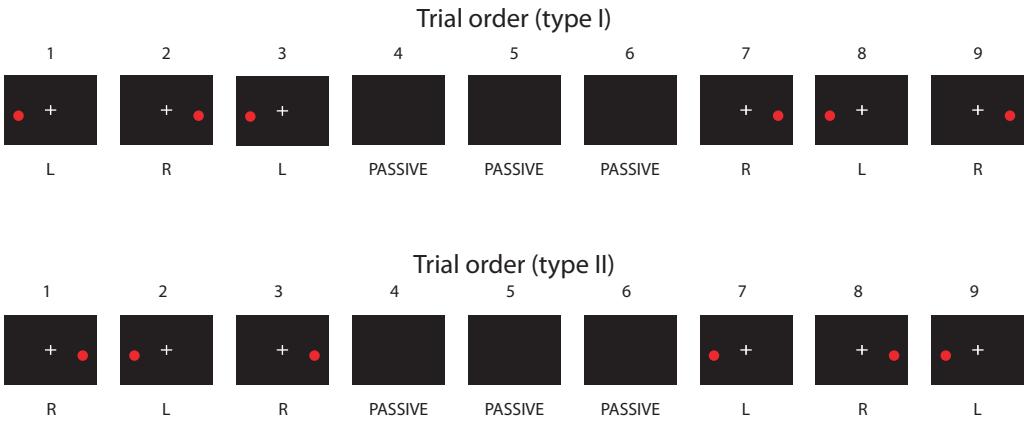

Figure 19: The presentation order of the trials (Table 1) was carefully randomized for each subject, but the three conditions (attend to left, attend to right, and passive) were locked to either trial order type I or II. Trials requiring attention to the left (L) or right (R) alternated so that two similar conditions never followed each other. Passive conditions were placed in the middle of the experiment.

7.4 MEG Acquisition

The MEG responses were recorded with a 306-channel (204 gradiometers and 102 magnetometers) whole-head magnetometer (Vectorview, Elekta Neuromag, Finland) in a three-layer magnetically shielded room at Low Temperature Laboratory, Aalto University School of Science and Technology. Before the actual MEG measurement, three electrooculogram (EOG) electrodes and four head position indicator (HPI) coils were attached to the subject's cheek, forehead and behind the ears (Figures 20 and 21). The EOG electrodes were used to detect vertical eye movements in order to remove eye blink artifacts from the MEG signal. The HPI coils were used to control the position of the head within the MEG measuring unit during the measurement.

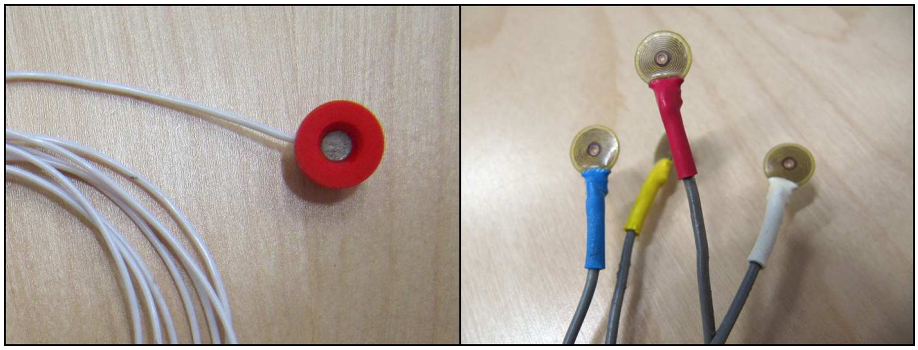

Figure 20: An electrooculogram (EOG) electrode (left) and head position indicator (HPI) coils (right).

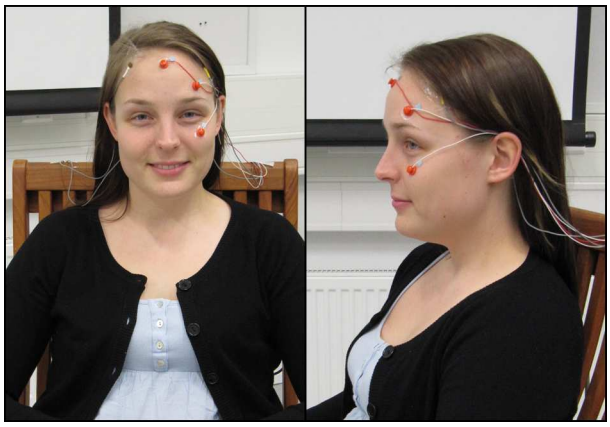

Figure 21: Three EOG electrodes and four HPI coils were attached to the subject's cheek, forehead and behind the ears to detect eye blinks and to control the position of the head within the helmet-shaped MEG sensor array.

It was ensured that at this point the subject was not wearing or carrying any metal objects. When the electrodes and coils were in place, the subject was taken inside the magnetically shielded room and seated under the MEG unit in a comfortable position. The head was positioned within the helmet-shaped MEG sensor array so that it touched lightly the surface of the helmet.

7.5 MEG Analysis

Only auditory-evoked responses were analyzed in the present study. Each auditory stimulus (adaptor and probe) was repeated 150 times for each condition and adaptor location (in total 9 trials, Table 1). The MEG signal was recorded for each repeated stimulus within a fixed time window (epoch) time-locked to 100 ms before and 500 ms after the stimulus onset. The recorded data within the epochs was averaged online. Vertical eye blink artifacts were measured with the EOG electrodes and epochs with deviations exceeding 150 μV were automatically discarded from the averaged data. The sampling rate was 1000 Hz with a bandwidth of 0.1–330 Hz. The averaged responses were filtered at 1–30 Hz and baseline corrected with respect to the 100-ms prestimulus period.

Forty-four gradiometer channels were picked from each temporal lobe for further analysis (Figure 22). The chosen channels were further analyzed in Matlab (The MathWorks, Inc.). The 20-ms latency caused by the audio system to the stimulus onset was corrected before the analysis. To average the data in channel pairs, vector sums were taken of each gradiometer channel pair, resulting in 22 vector sums for each hemisphere (Figure 23). The N1m peaks were then searched between 80 ms and 120 ms in each vector sum.

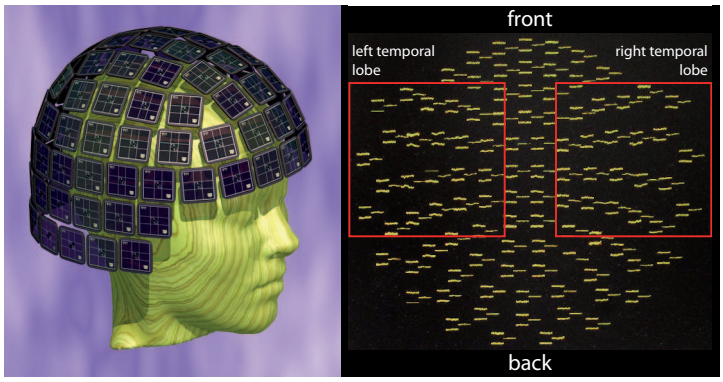

Figure 22: For each temporal lobe, 44 gradiometer channels (22 channel pairs) were recorded simultaneously with the MEG sensor array (left) and picked for further analysis. The 88 gradiometer channels of interest are marked approximately with the red rectangles. Image on the left courtesy of Elekta.

Prominent N1m peaks were found only in the vector sums of the right temporal lobe channels. For most subjects (83.3%), the N1m peaks recorded from the left temporal lobe either were so weak that they could not be reliably extracted from the data, or the N1m peak could not be distinguished from other peaks occurring around 100 ms. Therefore, the vector sums were further analyzed for the right temporal lobe only.

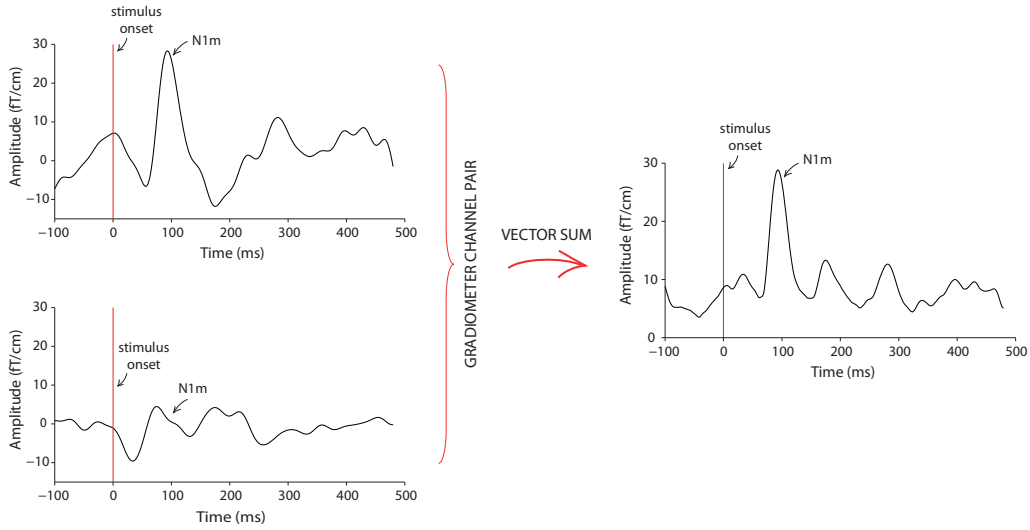

Figure 23: The vector sum of one gradiometer channel pair (left) is presented on the right. The 20-ms latency caused by the audio system to the stimulus onset was taken into account before computing the vector sums. In the vector sum, the amplitude of the N1m peak is approximately 30 fT/cm, and it occurs with approximately 100 ms latency.

For each subject, three best vector sums (vector sums with the largest N1m peaks) were picked from the right temporal lobe. The selected vector sums were averaged for each auditory stimulus (the adaptors and the probe) and trial to represent the subject's average N1m responses of the trials. The amplitudes of the N1m peaks and their latencies were saved for further analysis.

The location of the largest N1m amplitude peaks in the right temporal lobe did not vary much between the subjects. For 83.3% of the subjects the largest peaks were all among the four most popular vector sums. For the rest of the subjects (16.7%), one of the chosen channel pairs was among the four most popular channel pairs.

7.6 Statistical Analysis

The N1m peak amplitudes and latencies were statistically analyzed using a 3-way repeated-measures analysis of variance (ANOVA). The following dependent factors were used: sound type (adaptor, probe), condition (attend to left, attend to right, passive) and adaptor location ($-20°$, $0°$, $20°$). In addition, Newman-Keuls post hoc tests were performed when deemed necessary.

8 Results

8.1 Behavioral Performance

The visual task was difficult (mean hit rate 68.0%). Three of the subjects were only able to correctly identify a little less than half of the presented targets, and only two of the subjects identified them with more than 80% accuracy. However, none of the subjects were excluded due to poor behavioral performance, because the proportion of correct hits was clearly higher than that of false alarms, suggesting that the subjects had understood the task. In addition, 41.7% of the subjects systematically mixed blue with purple due to the proximity of the colors, which caused 68.5% of all false alarms.

8.2 MEG Results

Due to poor SNR, it was not possible to reliably determine the auditory-evoked N1m peaks recorded from the left hemisphere. In contrast, distinguishable N1m response amplitudes were found for all subjects in the right hemisphere for both the adaptor and the probe in all conditions and all adaptor locations. The amplitude of the N1m was significantly stronger for the probe ($\overline{x}$ = 23.9 fT/cm) than for the adaptor ($\overline{x}$ = 17.9 fT/cm; F[1,11] = 20.5, $p < 0.001$).

8.2.1 The Effect of Adaptor Location

The amplitude of the N1m response varied strongly with the adaptor location (F[2,22] = 17.0, $p < 0.0001$), as seen in Figure 24 and Figure 27. This effect was stronger for the probe than for the adaptor. When the adaptor was moved from $-20°$ to $0°$, the N1m responses to the probe increased from 16.3 fT/cm to 25.5 fT/cm ($p < 0.001$). When the adaptor was moved even farther from the probe, from $0°$ to $20°$, the amplitude of the N1m response to the probe increased further to 29.8 fT/cm ($p < 0.01$).

In contrast, the amplitude of the N1m response to the adaptor did not vary with the adaptor location as strongly as it did for the probe. Moving the adaptor from $-20°$ to $0°$ (amplitudes 16.3 fT/cm and 17.3 fT/cm, respectively) did not elicit a significantly larger N1m response amplitude (p = n.s.). However, a slightly significant increase was detected when the adaptor was moved further to $20°$, as the amplitude of the N1m response to the adaptor increased to 20.2 fT/cm ($p < 0.05$). As expected, both the probe and the adaptor elicited equally large N1m amplitudes at the adaptor location $-20°$ (16.3 fT/cm), because they were identical.

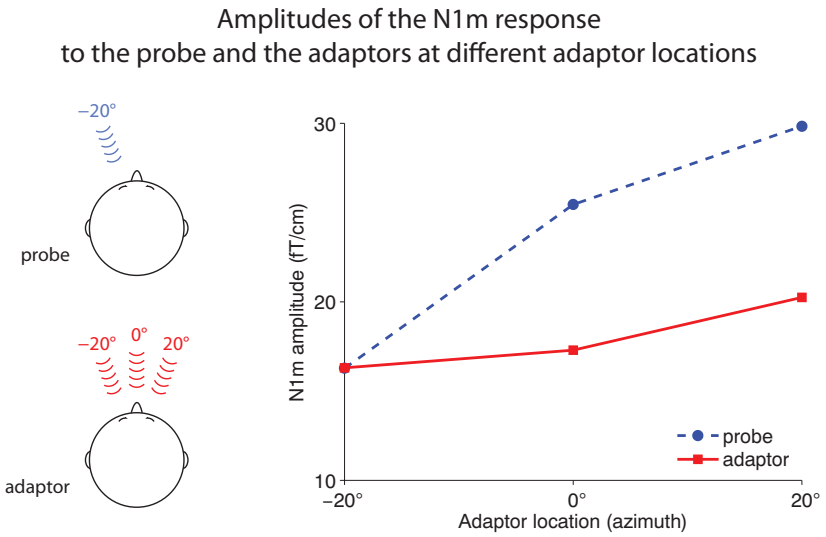

Figure 24: Left: The probe was always presented at the same location (−20°), while the adaptor had three possible locations (−20°, 0°, or 20°). Right: The N1m response amplitudes for the probe (dashed line) and the adaptors (solid line) at different adaptor locations. The amplitude of the N1m response to the probe increased systematically as the spatial separation between the adaptor and the probe increased ($p < 0.0001$). The amplitude of the N1m response to the adaptor showed a similar but a more moderate pattern. The increase in the N1m amplitude to the adaptor was only slightly significant when the adaptor was moved from 0° to 20° ($p < 0.05$), failing to be significant for the other two adaptor locations.

8.2.2 The Effect of Visual Attention on the Responses to the Probe

The effect of attention on the responses to the probe depended on the location of the adaptor. When the adaptor was presented at 20°, attending to visual stimuli presented either on the left or right yielded significantly larger response amplitudes (31.3 fT/cm and 32.0 fT/cm, respectively) compared to the passive condition (26.2 fT/cm, $p < 0.002$). Interestingly, the amplitude of the N1m was modulated equally by both attention conditions (attend to left and attend to right). In other words, attending visually toward (left) or away (right) from the probe equally increased the amplitude of the N1m response to the probe. This happened even though the probe was always presented in the left side and was, thus, outside the spatial focus of attention when visual attention was directed to the right (Figure 29).

However, when the adaptor was presented at −20° or 0°, attending to visual stimuli presented either on the left or right did not significantly modulate the N1m response to the probe compared to the passive condition ($F[4,44] = 2.4$, $p =$ n.s.). The amplitude of the N1m response to the probe did not vary significantly between the three different conditions (attend to left, attend to right, and passive) when the adaptor was located at −20° (15.6–17.1 fT/cm) or at 0° (23.2–26.8 fT/cm) (Figure 25).

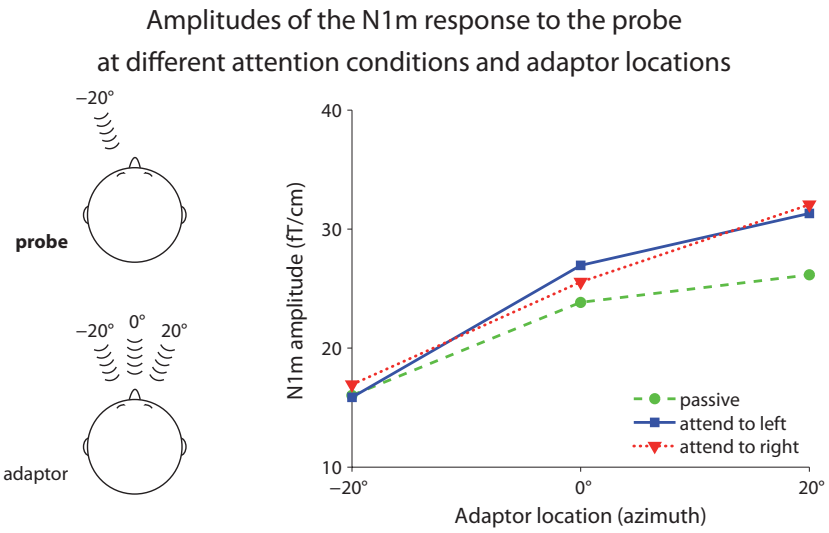

Figure 25: Left: The probe was always presented at the same location (−20°), while the adaptor had three possible locations (−20°, 0°, or 20°). Right: The amplitudes of N1m response to the probe in different conditions (passive, attend to left, and attend to right) and adaptor locations. When the adaptor was located at 20°, conditions requiring visual attention to the left or right evoked significantly larger amplitudes than the passive condition ($p < 0.002$). For other adaptor locations (−20° and 0°) the attention conditions did not have a significant effect to the amplitude. In all three conditions, the amplitude of the N1m response to the probe increased as the adaptor was moved farther away from the probe.

8.2.3 The Effect of Visual Attention on the Responses to the Adaptors

The amplitudes of the N1m responses to the adaptors were not the primary interest of this experiment, because the experiment was not designed to capture responses to sounds presented in the right hemifield. Attending to the visual stimuli presented either on the left or right did not have a significant effect on the amplitude of the N1m responses in any of the three conditions (attend to left, attend to right, and passive) at any adaptor angles ($F[4,44] = 2.4$, $p =$ n.s.), as can be seen in Figures 26 and 28.

8.2.4 The Effect of Visual Attention on the N1m Peak Latencies

The mean latency of the N1m peak was 101 ms for both the probe and the adaptor. The latency did not vary systematically depending on the sound type, condition, or adaptor location. The latency of the N1m peak did not depend on, e.g., the condition when the location of the adaptor was altered ($F[4,44] = 1.2$, p $<$ n.s.).

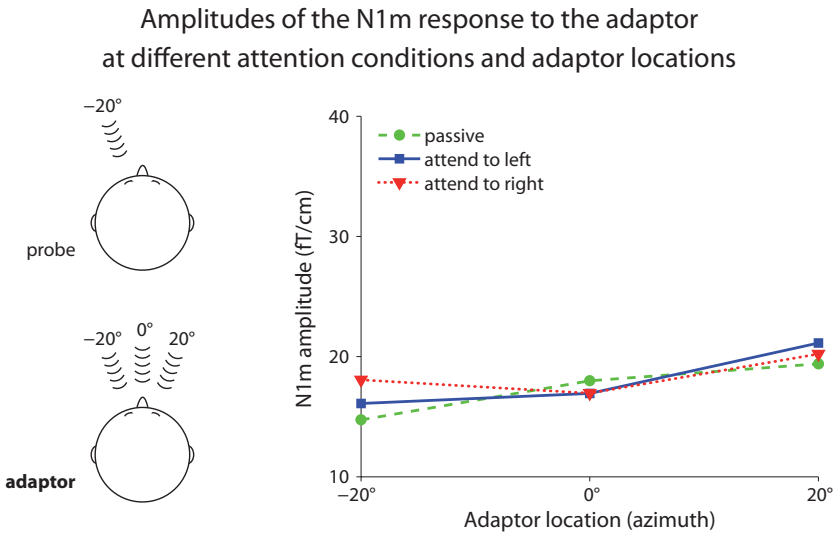

Figure 26: Left: The probe was always presented at the same location (−20°), while the adaptor had three possible locations (−20°, 0°, or 20°). Right: The amplitudes of N1m response to the adaptor at different locations and in different conditions (passive, attend to left, and attend to right). Attending to the left or right did not have an effect on the N1m responses compared to the passive condition.

9 Discussion

The purpose of this thesis was to investigate how spatially attending to visual stimuli affects auditory spatial processing in human auditory cortex. In a recent study it was found that the spatially selective human cortical neurons code the auditory space (sound source locations) with the relative activity of two opponent neuronal populations: one tuned to the left and the other to the right hemifield [2]. In addition, it was proposed that the majority of the neurons within one hemisphere are tuned to contralateral sound locations. In the present study, a similar experimental paradigm based on stimulus-specific adaptation was used to investigate whether this neural coding can be modulated by visual spatial attention.

In this study, the subjects were presented with unsynchronized streams of visual and auditory stimuli, and the brain responses to the auditory stimuli were recorded with the MEG. The impact of visual attention was measured under two different conditions: attend to left, and attend to right. In both conditions, the subject was instructed to attend to the visual stimuli (by doing a 1-back task) on the appropriate side. In addition, a third, passive condition (no attention), where the subject was silently reading a book of their choice, served as a baseline condition for the study. In all of the three conditions, the subject was instructed to ignore the sounds. It was found that the amplitude of the N1m response to the probe depended on two distinct factors: 1) visual spatial attention and 2) the separation between the probe and the adaptor.

9.1 The Spatial Selectivity of Auditory Cortical Neurons

A typical auditory-evoked field (AEF) with an N1m response peaking systematically around 100 ms after stimulus onset was found for most subjects, suggesting that the obtained MEG data represented the brain responses to the auditory stimuli as intended. As expected, the N1m responses measured from the left hemisphere were weaker than those recorded from the right side. The left-hemispheric data was, therefore, excluded from the results. It is very likely that the responses were weak mainly due to the fact that the majority of the sounds in the present study were presented in the left hemifield. Since the spatially selective auditory cortical neurons mostly consist of contralaterally tuned neurons, ipsilateral sounds do not evoke large N1m responses [2]. In addition, it was expected that the probe would elicit large N1m responses in the right hemisphere. The present results are in line with this hypothesis. As the probe was always presented in the left (at $-20°$), it activated the dominating contralateral-favoring left-tuned neuronal population in the right hemisphere, which lead to large response amplitudes.

As illustrated in Figure 27, the N1m response amplitudes increased systematically for the probe in all three conditions (attend to left, attend to right, and passive) in the right hemisphere, as the adaptor was moved farther away from the probe. This finding is consistent with a previous study where the stimulus-specific adaptation paradigm was used to reveal the population rate code of auditory space [2]. According to that study, the increasing pattern of the N1m amplitude indicates the

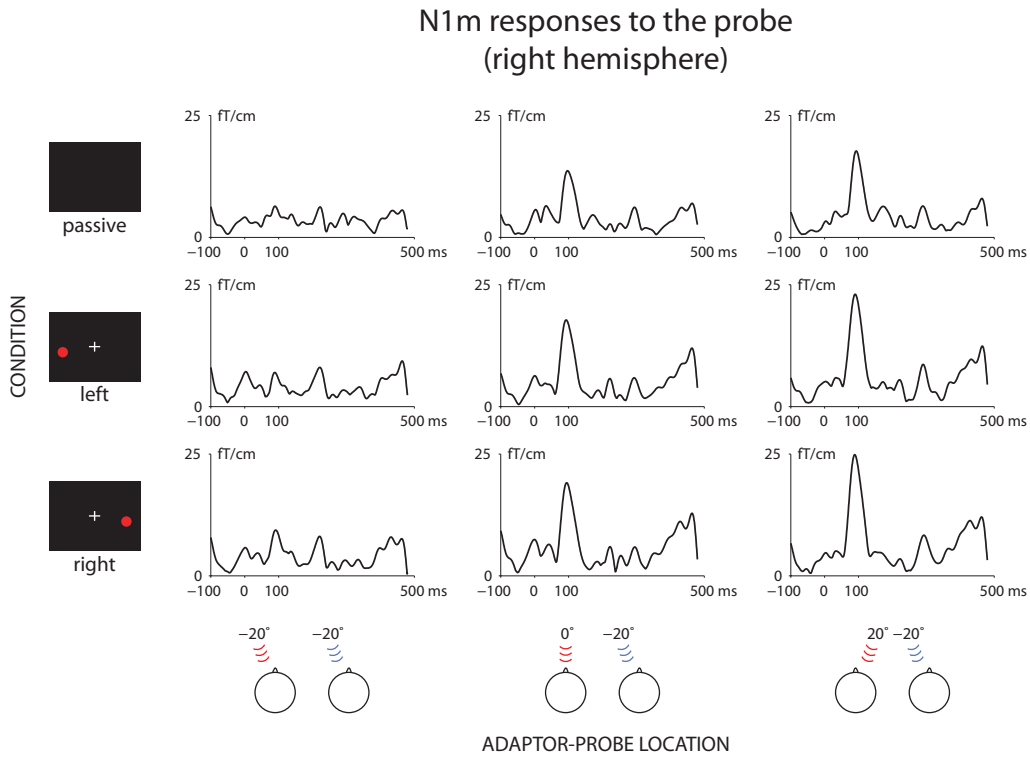

Figure 27: Grand-averaged event-related fields measured for the probe from the right hemisphere in different conditions. For the probe and the adaptor located at −20° and 20° (the third column), the N1m responses increased significantly in the two attention conditions compared to the passive condition. The effect of attention to the probe was equally prominent regardless of the direction of attention. The amplitudes increased systematically when the adaptor was moved farther away from the probe, an effect that has been proposed to indicate the existence of two opponent populations of neurons coding the auditory space. When the sounds were presented from the same location, they activated the same neuronal population and the adaptation was maximal (the first column). In contrast, when the sounds were presented in different hemifields, they activated different populations and no or very little adaptation occurred (the third column).

existence of the two neuronal populations, one tuned to sounds coming from the left and the other from the right hemifield. In addition, the population favoring contralateral sounds was found to dominate in each hemisphere. Based on the results of the previous study, the present results could be interpreted as follows: Two sounds that are presented in the same location activate the same neuronal population in the right hemisphere, leading to a maximal attenuation of the N1m response (the first column in Figure 27). In contrast, when the probe is presented in the left and the adaptor in the right hemifield, they activate different neuronal populations in the right hemisphere (mainly left-tuned and right-tuned populations, respectively). The adaptor does not, therefore, maximally activate the probe-preferring left-tuned population, which is why adaptation in the responses to the probe does not occur.

Thus, the attenuation of the N1m response is minimal, leading to large amplitudes of the response (the third column in Figure 27).

As expected, the brain responses in the right hemisphere were weaker to the adaptor than to the probe. Based on similar results reported in a previous study [2], it is justified to believe that the adaptors that might have activated the dominating contralaterally tuned neuronal population in the right hemisphere (at locations $-20°$ and $0°$, the first and the second column in Figure 28) failed to elicit large response amplitudes, because they were presented sequentially with the probe (located at $-20°$). The proximity of these two sounds caused nearly maximal adaptation in the responses of the adaptors, and the responses remained small. In contrast, for the third adaptor location ($20°$, the third column in Figure 28), no or only little adaptation occurred, because the sounds were presented in opposite hemifields. However, the increase in the amplitude was not as large as for the probe with a similar adaptor-probe angle combination (the third column in Figure 27), because the adaptor was presented in the same side (right) compared to the hemisphere that was being measured. Ipsilateral sounds do not activate the auditory cortical neurons as strongly as contralateral sounds, as shown in an earlier study [1, 2].

9.2 The Effects of Attention

The main interest of this thesis was to investigate whether the spatial processing in the human auditory cortex can be modulated by attending to visual stimuli. The principal finding of this study was that attending to visual stimuli indeed increased the amplitude of the N1m response elicited by the probe sound compared to the passive condition during which the subject was reading a book (Figure 29). This finding suggests that the increase in the amplitude was somehow related to attending to the visual stimuli. However, against all expectations, the amplitude of the N1m response to the probe increased equally in both attention conditions (attend to left, and attend to right) regardless of whether the attended visual stimuli were presented in the same or in the opposite side as the probe. This is a surprising finding, because the probe was outside the focus of visual spatial attention half of the time, i.e., when attention was directed to the opposite (right) side. The effect was visible and significant especially when the probe and the adaptor were presented in the opposite hemifields, although a similar increase, albeit smaller and not statistically significant, could also be seen when the adaptor was presented directly in front of the subject (Figure 25, adaptor azimuth $0°$). However, no attentional effect could be observed when the adaptor was presented in the same location with the probe (Figure 25, adaptor azimuth $-20°$). The results indicate that the effect of visual attention increases when the separation of the probe and adaptor grows.

In the light of previous studies, it does not seem plausible that directing visual attention away from the probe (to the right side) would produce equally prominent N1m responses to the probe as directing visual attention toward the probe (to the left side). Several studies concerning cross-modal spatial attention have reported that visual attention increased the brain responses to task-irrelevant auditory stimuli only

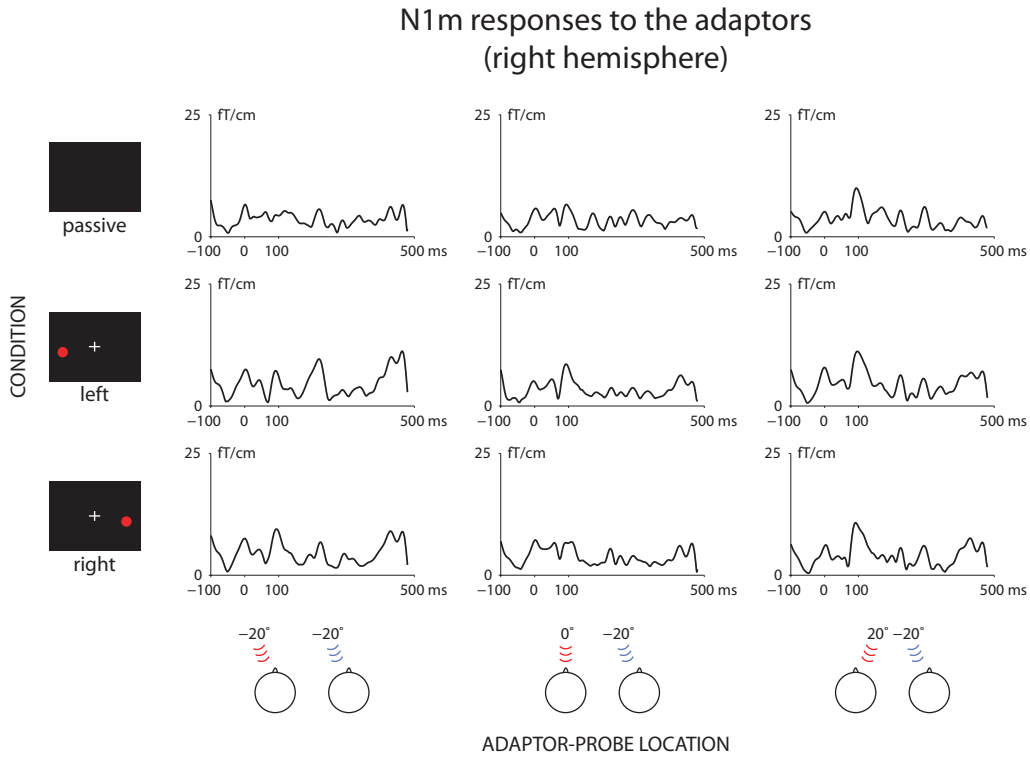

Figure 28: Grand-averaged event-related fields measured for the adaptors from the right hemisphere in different conditions. The adaptors did not elicit large N1m responses. The adaptation caused by the probe diminished the activation for the adaptors presented at $-20°$ and $0°$, whereas the third adaptor was presented at an ipsilateral location compared to the currently measured hemisphere and, thus, was not able to maximally activate the neurons in the right hemisphere.

when the visual target stimuli were spatially coincident with the auditory stimuli (i.e., presented in the same location) [8, 7]. Conversely, in the same studies it was found that when visual attention was directed away from the auditory stimuli, the brain responses to the auditory stimuli decreased. The present results are only partly in line with the previous studies, that is, when the attention was visually directed to the side of the probe (attend to left). In this condition, an increase in the amplitude of the N1m response to the probe could be seen, as reported in previous studies. However, the present results contradict with the previous findings about that visually attending to the opposite side would decrease the amplitudes in the unattended modality. In the present study, attention increased the N1m response to the probe even when the side where the probe was presented was visually unattended.

There are some factors that could have had an effect on the results of the present study. First, similar effects to the brain responses of unattended auditory stimuli have been found in previous experiments considering multisensory objects. In a previous study it was found that attention can spread cross-modally, forming multisensory objects consisting of features from different modalities [45]. It was proposed

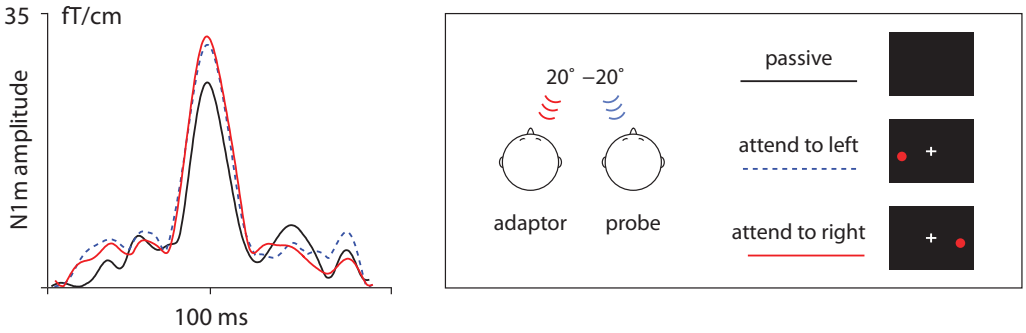

Figure 29: The effect of attention on the N1m response amplitudes in different conditions for the probe. Attending visually to the left or right equally increased the N1m responses compared to the passive condition, when the probe and the adaptor were in their most lateral positions (probe: −20°, adaptor: 20°). The effect of visual attention to the N1m response amplitudes was present even when attention was directed away (attend to right) from the location of the probe.

that attention that spreads this way may modulate the brain responses to unattended features of the multisensory object, even if these features were spatially disparate. However, in studies concerning this phenomena, auditory and visual stimuli were typically presented synchronized in time, which was probably enough to create an illusion of an multisensory object, albeit consisting of spatially disparate features. In contrast, the auditory stimuli in the present study were not related with the visual stimuli in time or any other meaningful way, other than that in the attend to left condition the auditory stimuli were spatially congruent with the visual stimuli. The auditory stimuli in the present study were designed in such way that temporally coinciding occurrences with visual stimuli were rare (the presented sounds coincided with only 5% of all presented visual stimuli), to prevent the stimuli from forming multisensory objects. Therefore, it is not likely that the sounds occurring on one side would have been merged into a multisensory object with the visual stimuli of the opposite side, and, thus, caused equally large N1m responses regardless of the direction of visual attention.

Second, it is well known that the ability to localize sounds can be biased toward synchronously presented visual stimuli if the visual stimuli are presented not more than 20° from the auditory stimuli (ventriloquism effect) [43, 44]. In the present study, the auditory and the visual stimuli were not synchronous. In addition, when visual attention was directed to the opposite side as the probe (attend to right), the distance between the visual and auditory stimuli (40°) exceeded the limit set for the ventriloquism effect (20°) in the previous studies. Thus, it is unlikely that such effect influenced the results even if it actually did affect asynchronous stimuli as well.

Third, in a previous study it was found that short-term memory load of a visual task can have similar enhancing effects to the auditory N1m as what was found in the present study [46]. If the modulation of the auditory N1m was caused purely by the short-term memory load of the visual task, the increase of the N1m should be visible

at all adaptor angles in the responses of the adaptor and the probe when the visual 1-back task was present, i.e., in the conditions involving visual attention. However, no attention-related increase in the N1m amplitudes of the probe could be seen for the adaptor location $-20°$ compared to the passive condition (Figure 25). Similarly, attending to the left or right did not enhance the N1m responses to the adaptor for any adaptor location (Figure 26). Thus, even if the visual task itself had some influence on the N1m response, something else is modulating the response even more and covering the memory load effects of the visual task. Also opposite results have been found where the visual task decreased the responses to the irrelevant auditory stimuli [53], but no such decrease was observed in the present results compared to the passive condition.

Finally, it is worth considering that overt attention, involving shifting the gaze toward the attended visual stimuli, was used in the present study instead of covert attention (mentally focusing on a location in visual space, without moving the eyes). Let us imagine that overtly attending to a visual stimulus presented either in the left or right would shift that stimulus into the spatial center of the visual attentional space, leaving the auditory attentional space untouched. This would lead to different coordinate systems between the auditory and visual modalities; the auditory modality would be head-centered, whereas the visual modality would depend on the movements of the eyes, being eye-centered. If this was true, the probe sound would always be presented in its original location ($-20°$), because it is not possible to turn the head within the MEG measuring unit, while the visual stimuli would be presented at $0°$ due to the eye movement toward the visual stimuli. This would result in equally large N1m responses to the probe in both conditions involving attention, because there would be no difference between attending to the left and attending to the right. Moreover, the separation between the probe sound and the visual stimuli would be a constant $20°$ (the upper limit for the ventriloquism effect), making mislocalizing the sounds toward the visual stimuli possible and, thus, leading to enhanced N1m responses. However, as stated before, it is unlikely that the ventriloquism effect influenced the results, because the visual and the auditory stimuli of the present study were unsynchronized.

10 Conclusions

The results of this thesis implicate that the neural coding of the auditory space can be modulated by attending to visual stimuli. Visual attention was found to increase the overall sensitivity of spatially selective auditory cortical neurons regardless of the direction of the visual attention. The effect of visual attention on this neural coding appeared to be dependent on the spatial separation between the probe and the adaptor. No effect was observed when the probe and the adaptor were presented at the same location, whereas the effect was the largest when the two sounds were presented in opposite hemifields (Figure 25).

It is likely that the brain responses to the auditory stimuli were mainly modulated by visual attention and not, for example, by multisensory object forming or by the short-term memory load caused by the visual task. However, it is also possible that overtly attending to the visual stimuli shifted the eye-centered coordinates in relation to the head-centered coordinates, making no difference between the attended locations and contributing, thus, to the enhanced N1m responses in both conditions. Further experiments will be required in order to tell whether overt and covert attention would have produced different results.

It is difficult to say why the present results contradict with some previous studies [8, 7] where unattended locations did not receive enhanced neuronal processing. However, the differences in the results could be due to the fact that different brain imaging methods (EEG and MEG) were used in different studies. On the one hand, the EEG is not spatially as accurate as the MEG, but on the other hand, the MEG does not measure the activity of all the neurons that the EEG is capable of measuring.

Finally, the population rate code investigated in the present study is suitable for modeling the spatial coding of auditory space in the horizontal plane only. Little is known about the neural basis of auditory spatial coding in the vertical direction and, thus, the effects of visual attention on the neural coding of the full auditory space remain unknown.

References

[1] Werner-Reiss, U., and Groh, J. M. A Rate Code for Sound Azimuth in Monkey Auditory Cortex: Implications for Human Neuroimaging Studies. *The Journal of Neuroscience,* 2008, vol. 28, p. 3747–3758.

[2] Salminen, N. H., May, P. J. C., Alku, P., and Tiitinen, H. A Population Rate Code of Auditory Space in the Human Cortex. *PLoS ONE,* 2009, vol. 4, e7600.

[3] Butler, R. A. The influence of spatial separation of sound sources on the auditory evoked response. *Neuropsychologia,* 1972, vol. 10, p. 219–225.

[4] Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. Electrical Signs of Selective Attention in the Human Brain. *Science,* 1973, vol. 182, p. 177–180.

[5] Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., and Bloom, F. E. Modulation of early sensory processing in human auditory cortex during auditory selective attention. *The Proceedings of the National Academy of Sciences (US),* 1993, vol. 90, p. 8722–8726.

[6] Teder-Sälejärvi, T., and Hillyard, S. A. The gradient of spatial auditory attention in free field: An event-related potential study. *Perception & Psychophysics,* 1998, vol. 60, p. 1228–1242.

[7] Eimer, M. Crossmodal links in spatial attention between vision, audition, and touch: evidence from event-related brain potentials. *Neuropsychologia,* 2001, vol. 39, p. 1292–1303.

[8] Hillyard, S. A., Simpson, G. V., Woods, D. L., Van Voorhis, S., and Munte, T. F. Event-related brain potentials and selective attention to different modalities. In: Reinoso-Suarez, F., Ajmone-Marsan, C. (editors). *Cortical integration.* New York, Raven Press, 1984, p. 395–414.

[9] Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., and Grady, C. L. The Functional Organization of Human Extrastriate Cortex: A PET-rCBF Study of Selective Attention to Faces and Locations. *The Journal of Neuroscience,* 1994, vol. 14, p. 6336–6353.

[10] Shomstein, S., and Yantis, S. Control of Attention Shifts between Vision and Audition in Human Cortex. *The Journal of Neuroscience,* 2004, vol. 24, p. 10702–10706.

[11] Moore, B. C. J. *An Introduction to the Psychology of Hearing.* 5th ed. London, Academic Press, 2006.

[12] Middlebrooks, J. C., and Green, D. M. Sound Localization by Human Listeners. *Annual Review of Psychology,* 1991, vol. 42, p. 135–159.

[13] Goldstein, E. B. *Sensation and Perception.* 6th ed. Pacific Grove, CA, Wadsworth, 2002.

[14] Macpherson, E. A., and Middlebrooks, J. C. Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America,* 2002, vol. 111, p. 2219–2236.

[15] Gardner, B., and Martin, K. HRTF Measurements of a KEMAR Dummy-Head Microphone. *MIT Media Lab Perceptual Computing – Technical Report #280.* Cambridge, MA, MIT Media Lab, May 1994.

[16] Wightman, F. L., and Kistler, D. J. Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America,* 1989, vol. 85, p. 858–867.

[17] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America,* 1993, vol. 94, p. 111–123.

[18] Gardner, W. G. *3-D Audio Using Loudspeakers.* 1st ed. Norwell, MA, Kluwer Academic Publishers, 1998.

[19] Baars, B. J., and Gage, N. M. *Cognition, Brain, and Consciousness: An Introduction to Cognitive Neuroscience.* 1st ed. London, Academic Press, 2007.

[20] Kuzniecky, R. I., and Jackson, G. D. *Magnetic Resonance in Epilepsy: Neuroimaging Techniques.* 2nd ed. Burlington, MA, Academic Press, 2005.

[21] Rodden, F. A., and Stemmer, B. A Brief Introduction to Common Neuroimaging Techniques. In: Stemmer, B., and Whitaker, H. A. (editors). *Handbook of the Neuroscience of Language.* 1st ed. London, Academic Press, 2008, p. 57–67.

[22] Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. *Cognitive Neuroscience: The Biology of the Mind.* 2nd ed. New York, NY, W. W. Norton & Company, inc., 2002.

[23] Neukirch, M., Hegerl, U, Kötitzt, R., Dorn, H., Gallinat, U., and Herrmann, W. M. Comparison of the Amplitude/Intensity Function of the Auditory Evoked N1m and N1 Components. *Neuropsychobiology,* 2002, vol. 45, p. 41–48.

[24] Lü, Z., Williamson, S. J., Kaufman, L. Human auditory primary and association cortex have differing lifetimes for activation traces. *Brain Research,* 1992, vol. 572, p. 236–241.

[25] Woods, T. M., Lopez, S. E., Long, J. H., Rahman, J. E., and Recanzone, G. H. Effects of Stimulus Azimuth and Intensity on the Single-Neuron Activity in the Auditory Cortex of the Alert Macaque Monkey. *Journal of Neurophysiology,* 2006, vol. 96, p. 3323–3337.

[26] Recanzone, G. H. Spatial processing in the auditory cortex of the macaque monkey. *The Proceedings of the National Academy of Sciences (US),* 2000, vol. 97, p. 11829–11835.

[27] Butler, R. A. Effect of Changes in Stimulus Frequency and Intensity on Habituation of the Human Vertex Potential. *The Journal of the Acoustical Society of America,* 1968, vol. 43, p. 945–950.

[28] Spitzer, H., Desimone, R., and Moran, J. Increased Attention Enhances Both Behavioral and Neuronal Performance. *Science,* 1988, vol. 240, p. 338–340.

[29] Gandhi, S. P., Heeger, D. J., and Boynton, G. M. Spatial attention affects brain activity in human primary visual cortex. *The Proceedings of the National Academy of Sciences (US),* 1999, vol. 96, p. 3341–3319.

[30] Posner, M. I. *Cognitive Neuroscience of Attention.* 1st ed. New York, The Guilford Press, 2004.

[31] Näätänen, R., Ilmoniemi, R. J., and Alho, K. Magnetoencephalography in Studies of Attention. In: Lang, P. J., Simons, R. F., and Balaban, M. (editors). *Attention and Orienting: Sensory and Motivational Processes.* New Jersey, Lawrence Erlbaum Associates, 1997, p. 307–326.

[32] Mueller, V., Brehmer, Y., von Oertzen, T, Li, S., and Lindenberger, U. Electrophysiological correlates of selective attention: A lifespan comparison. *BMC Neuroscience,* 2008, vol. 9.

[33] Picton, T. W., and Hillyard, S. A. Human auditory evoked potentials. II: Effects of attention. *Electroencephalography and Clinical Neurophysiology,* 1974, vol. 36, p. 191–199.

[34] Woldorff, M. G., and Hillyard, S. A. Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology,* 1991, vol. 79, p. 170–191.

[35] Fujiwara, N., Nagamine, T., Imai, M., Tanaka, T., and Shibasaki, H. Role of the primary auditory cortex in auditory selective attention studied by whole-head neuromagnetometer. *Science,* 1998, vol. 7, p. 99–109.

[36] Jäncke, L., Mirzazade, S., and Shah, N. J. Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neuroscience Letters,* 1999, vol. 266, p. 125–128.

[37] Rinne, T., Pekkola, J., Degerman, A., Autti, T., Jääskeläinen, I. P., Sams, M., and Alho, K. Modulation of Auditory Cortex Activation by Sound Presentation Rate and Attention. *Human Brain Mapping,* 2005, vol. 26, p. 94–99.

[38] Grady, C. L., Van Meter, J. W., Maisog, J. M., Pietrini, P., Krasuski, J., and Rauschecker, J. P. Attention-related modulation of activity in primary and secondary auditory cortex. *Cognitive Neuroscience and Neuropsychology,* 1997, vol. 8, p. 2511–2516.

[39] Alho, K., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Tervaniemi, M., Reinikainen, K., Zeffiro, T., and Näätänen, R. Selective tuning of the left and right auditory cortices during spatially directed attention. *Cognitive Brain Research,* 1999, vol. 7, p. 335–341.

[40] O'Leary, D. S. et al. Auditory and Visual Attention Assessed With PET. *Human Brain Mapping,* 1997, vol. 5, p. 422–436.

[41] Schwent, V. L., and Hillyard, S. A. Evoked potential correlates of selective attention with multi-channel auditory inputs. *Electroencephalography and Clinical Neurophysiology,* 1975, vol. 38, p. 131–138.

[42] Schwent, V. L., Snyder, E., and Hillyard, S. A. Auditory evoked potentials during multichannel selective listening: Role of pitch and localization cues. *Journal of Experimental Psychology: Human Perception and Performance,* 1976, vol. 2, p. 313–325.

[43] Bertelson, P., Vroomen, J., de Gelder, B., and Driver, J. The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics,* 2000, vol. 62, p. 321–332.

[44] Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., and Schirillo, J. A. Visual Localization Ability Influences Cross-Modal Bias. *Journal of Cognitive Neuroscience,* 2003, vol. 15, p. 20–29.

[45] Busse, L., Roberts, K. C., Crist, E. R., Weissman, D. H., and Woldorff, M. G. The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences,* 2005, vol. 102, p. 18751–18756.

[46] Valtonen, J., May, P., Mäkinen, V., and Tiitinen, H. Visual short-term memory load affects sensory processing of irrelevant sounds in human auditory cortex. *Cognitive Brain Research,* 2003, vol. 17, p. 358–367.

[47] Shulman, G. L. et al. Top-down Modulation of Early Sensory Cortex. *Cerebral Cortex,* 1997, vol. 7, p. 193–206.

[48] Rees, G., Frith, C., and Lavie, N. Processing of irrelevant visual motion during performance of an auditory attention task. *Neuropsychologia,* 2001, vol. 39, p. 937–949.

[49] Knudsen, E. I. Fundamental Components of Attention. *The Annual Review of Neuroscience,* 2007, vol. 30, p. 57–78.

[50] Palomäki, K. J., Tiitinen, H., Mäkinen, V., May, P. J. C., and Alku, P. Spatial processing in human auditory cortex: The effects of 3D, ITD, and ILD stimulation techniques. *Cognitive Brain Research,* 2005, vol. 24, p. 364–379.

[51] Salminen, N. H., Tiitinen, H., Yrttiaho, S., and May, P. J. C. The neural code for interaural time difference in human auditory cortex. *The Journal of the Acoustical Society of America,* 2010, vol. 127, p. EL60–EL65.

[52] Gardner, B., and Martin, K. HRTF Measurements of a KEMAR Dummy-Head Microphone. Online document. Updated July 18, 2000. Date cited Jan 20, 2010. Available from: `http://sound.media.mit.edu/resources/KEMAR.html`.

[53] Dyson, B. J., Alain, C., and He, Y. Effects of visual attentional load on low-level auditory scene analysis. *Cognitive, Affective, & Behavioral Neuroscience,* 2005, vol. 5, p. 319–338.